

ABSTRACT

Title of Dissertation: THE ALIGNMENT OF EIGHTH-GRADE
MATHEMATICS INSTRUCTION ACROSS
ACADEMIC TRACKS WITH STATEWIDE HIGH
STAKES TESTS: IMPLICATIONS FOR TEST
PERFORMANCE

Felice S. Shore, Doctor of Philosophy, 2003

Dissertation directed by: Professor Patricia Campbell
Department of Curriculum and Instruction

This study described the nature of the link between the types of instructional strategies employed in different academic tracks of eighth-grade mathematics classrooms and student achievement on a state performance assessment (PA). Survey data was obtained from 51 teachers in one school district who responded to a two-part questionnaire ascertaining teacher background characteristics and instructional strategies. A reform score was calculated that represented the percent of instructional time devoted to teaching strategies consistent with the focus of the PA. All data were analyzed at the class level, specifically examining any notable differences among tracks.

Across all teachers, variation in instructional strategies was modest. There were no significant differences between mean reform scores across three courses. Yet, for

Algebra II, a significant relationship was found between the amount of reform instruction and achievement on the PA. Overall, however, course level was not found to be a moderator between those two variables. Additionally, a model whereby course level acts as a mediator between reform instruction and student performance was not substantiated by the data.

There was an inequitable distribution of teachers with mathematics credentials in the surveyed classrooms, with lower-level courses being taught by teachers with lower certification levels. The finding that credentials influenced achievement above and beyond course level begs further research. Furthermore, how the influences of credentials and pressure from other high-stakes tests manifest themselves in the learning environment would be substantive topics for future studies.

An observational component of this study described teaching styles of three teachers with respect to instructional alignment with expectations implied in State Learning Outcomes. Two of the three teachers were judged to have content- and pedagogical content-knowledge deficiencies, limiting their ability to help students learn mathematics with understanding. The third teacher used pedagogical practices more likely to support the goal of students' learning meaningful mathematics with understanding. This study's qualitative component suggests further research examining teacher knowledge of mathematical content and pedagogy and its links to teacher practices and teacher questioning.

THE ALIGNMENT OF EIGHTH-GRADE MATHEMATICS INSTRUCTION
ACROSS ACADEMIC TRACKS WITH STATEWIDE HIGH STAKES TESTS:
IMPLICATIONS FOR TEST PERFORMANCE

By

Felice S. Shore

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2003

Advisory Committee:

Professor Patricia Campbell, Chair
Professor James Fey
Professor Anna Graeber
Professor Gregory Hancock
Professor Steven Selden

ACKNOWLEDGEMENTS

I wish to thank my husband, Gregg, without whose continued love and support this dissertation and degree would not have been possible. I would also like to thank our family, which grew at the same rate as this thesis. Sam, Eli, and Joshua were incredibly understanding and patient when mommy had to work behind closed doors.

I wish to thank members of my dissertation committee, James Fey, Anna Graeber, and Steve Selden for their thoughts and suggestions concerning this research project. A special thanks goes to Greg Hancock who aided me throughout this process, lending me his expertise, as well as a voice of calm, reason, and encouragement when I needed it most.

I owe tremendous gratitude to my advisor, Patricia Campbell, who always made time for me, accommodating my hectic schedule, and who pushed me along with just the right balance of encouragement and pressure.

I wish to commend all of the teachers who participated in this study, especially the three who allowed me into their classrooms for observations. I also would like to thank the District personnel in the Mathematics Office and the Research Office who spent time with me on many occasions to provide me information and data. They were supportive of my efforts in this project, and were extremely helpful and kind to me

Finally, I want to thank my friends Anne Wolf, Eileen D'Andrea, and Jenny Harbold who provided me both technical and emotional support throughout the writing.

TABLE OF CONTENTS

| | |
|---|-----|
| CHAPTER 1: INTRODUCTION..... | 1 |
| Research Context..... | 2 |
| Rationale..... | 10 |
| Research Questions..... | 19 |
| Classroom Observations..... | 19 |
| Key Terms..... | 20 |
| Significance and Purpose..... | 22 |
| Limitations..... | 22 |
| Assumptions..... | 23 |
| CHAPTER 2: REVIEW OF THE LITERATURE..... | 25 |
| High Stakes Testing..... | 25 |
| Performance Assessments as an Instruction for Educational Change..... | 28 |
| Barriers to Reform Instruction..... | 34 |
| Equity Issues in Assessment..... | 36 |
| Tracking and Opportunity to Learn..... | 38 |
| Effects of Reform Instruction on Achievement..... | 45 |
| Summary..... | 51 |
| CHAPTER 3: METHODS..... | 53 |
| Population and Study Site..... | 53 |
| Sample..... | 54 |
| Mathematics in Arbor County School District Middle Schools..... | 60 |
| Instruments and Variables..... | 65 |
| Data Analysis..... | 77 |
| CHAPTER 4: RESULTS..... | 84 |
| Descriptive Analyses..... | 84 |
| Instructional Models..... | 90 |
| Predicting Reform Instruction..... | 108 |

| | |
|---|-----|
| Achievement on the State Performance Assessment..... | 109 |
| Summary of Results..... | 114 |
| CHAPTER 5: OBSERVATIONS..... | 118 |
| Method..... | 118 |
| Results..... | 129 |
| Survey Scores and Observed Instruction..... | 238 |
| CHAPTER 6: CONCLUSIONS..... | 241 |
| Implications..... | 242 |
| Limitations..... | 251 |
| Recommendations..... | 254 |
| APPENDIX A: LETTER TO PRINCIPALS..... | 256 |
| APPENDIX B: SURVEY, PART 1..... | 258 |
| APPENDIX C: SURVEY, PART 2..... | 263 |
| APPENDIX D: LETTER TO PRINCIPALS FOR OBSDERVATIONS..... | 268 |
| APPENDIX E: OBSERVATION FORMS..... | 269 |
| APPENDIX F: POST-OBSERVATION NOTES..... | 270 |
| REFERENCES..... | 271 |

LIST OF TABLES

| | | |
|----------|---|-----|
| Table 1 | Sample Means (and Standard Deviations) for Middle School Demographic Characteristics for Schools in Analytic Sample or Excluded from Analytic Sample..... | 58 |
| Table 2 | Number of Sections Offered for Different Courses..... | 59 |
| Table 3 | Teacher Variables Assessed in the Questionnaire..... | 67 |
| Table 4 | Frequency of Instructional Strategy Use as Converted to Days in the School Year..... | 72 |
| Table 5 | Research Questions, Sources, and Analysis..... | 83 |
| Table 6 | Frequencies of Certification Types of Teachers in Three Levels of Courses..... | 85 |
| Table 7 | Teacher Background Characteristics for three course levels..... | 86 |
| Table 8 | Comparisons of Means for Classroom Characteristics..... | 88 |
| Table 9 | Differences in Content Coverage for Three Courses..... | 90 |
| Table 10 | Perceptions of Conflict by Course..... | 93 |
| Table 11 | Alpha-Factored Items from Questionnaire, Part 2 that Made up Accord Factor..... | 94 |
| Table 12 | Means and Standard Deviations for Accord Factor..... | 95 |
| Table 13 | Significant Correlations among Relevant Variables..... | 97 |
| Table 14 | Reform and Traditional Teaching Strategies..... | 100 |
| Table 15 | Percentage of Class Time Engaged in Reform Instructional Strategies..... | 101 |
| Table 16 | Frequency of Reported Reform or Traditional Instructional Strategies by Class Time..... | 103 |
| Table 17 | Summary of Hierarchical Regression Analysis for Variables Predicting Achievement on PA..... | 112 |

| | | |
|----------|--|-----|
| Table 18 | Teacher Credentials as a Significant Predictor for PA Achievement..... | 113 |
| Table 19 | Time Table for Observations of Ms. Drake’s Fundamentals Class..... | 131 |
| Table 20 | Time Table for Observations of Ms. Drake’s GT Class..... | 142 |
| Table 21 | Time Table for Observations of Ms. Henderson’s Pre-Algebra Class..... | 161 |
| Table 22 | Time Table for Observations of Ms. Henderson’s Algebra II Class..... | 170 |
| Table 23 | Time Table for Observations of Ms. Miller’s Algebra I Class..... | 189 |
| Table 24 | Time Table for Observations of Ms. Miller’s Algebra II (GT) Class..... | 211 |
| Table 25 | Reform Scores for Observed Teachers..... | 239 |

LIST OF FIGURES

| | | |
|----------|---|-----|
| Figure 1 | Mediational model, in which level of reform is the mediator..... | 81 |
| Figure 2 | Moderator model, in which course level is the moderator..... | 82 |
| Figure 3 | Traditional and reform instructional strategies that were among the reported top five most-used strategies in any one of the three courses..... | 104 |
| Figure 4 | Traditional and reform instructional strategies that were among the reported five least-used strategies in any one of the three courses..... | 106 |
| Figure 5 | Ms. Henderson’s classroom..... | 159 |
| Figure 6 | Arrangement of desks and students in Ms. Miller’s Algebra I class..... | 188 |

CHAPTER 1: INTRODUCTION

High-stakes testing has played an increasingly prominent role in our public school systems over the years. In recent years, high-stakes tests have been used not only for accountability purposes, but also as a way to help improve instruction. At the same time, trends in mathematics education reform have required teachers to be more flexible, resourceful, and more pedagogically and subject matter knowledgeable than had been previously expected. In order to effect an instructional change that sought to make mathematics more meaningful to students, states began instituting tests that purported to measure how well students were able to use the mathematics they learned to solve more in-depth, realistic, and context-bound problems. Such tests are typically referred to as “performance assessments.”

At the same time, most public school districts have continued to place students in different mathematics courses based on some measure of academic ability. Some researchers have posited that the differential instruction and content that students receive as a result of this “tracking” has the effect of widening the gap in student achievement.

One purpose of this study was to describe the nature of the link between the types of instructional strategies employed in different academic tracks of eighth-grade mathematics classrooms and student achievement on a state performance assessment. Another component of this study provides an in-depth look at the classroom strategies of three eighth-grade mathematics teachers.

This chapter presents the context in which this study was conducted from the perspective of current trends in mathematics education as well as the locale for the

study, followed by the rationale for the particular research questions addressed. A list of key terms is provided, as are the limitations and assumptions of this study.

Research Context

Standardized testing in education has had a long and often controversial history in the United States. Resnick (1981) maintains that school districts are expected to measure and report the state of schooling and student performance for reasons of efficiency, objectivity, and the identification of meritorious individuals. Furthermore, many policy makers are attracted to the notion that testing can be an inexpensive way to effect change in schools.

Popham (1987) claimed that well conceived and appropriately implemented assessments could in fact be used as vehicles to push instructional change. He called the notion “measurement-driven instruction” (MDI) and maintained that it “occurs when a high-stakes test of educational achievement, because of the important contingencies associated with the students’ performance, influences the instructional program that prepares students for the test” (p. 680). He further defined two kinds of high-stakes tests: those that have serious consequences for the individual test-takers, such as ones used to make promotional or instructional decisions, and those used to assess the quality of instruction in school systems, such as statewide tests whose results are reported by district.

Popham’s main argument is that high-stakes tests act as “curricular magnets” and thus those tests are used for “instructional clarification” (p. 681). Accordingly, it is crucial for those tests to be criterion-referenced, rather than norm-referenced, because of

the “descriptive clarity” (p. 680) intrinsic to criterion-referenced tests. That is, since teachers inevitably want their students to do well on such tests, they will focus instruction on the clearly defined content assessed by those tests. Although the proposal might seem reasonable, MDI is not without criticism. Bracey (1987) grants that although MDI is not an inherently flawed idea, in practicality, the negative outcomes are far too ubiquitous to ignore. He contends that MDI is actually at work whenever assessment programs “interrupt” the “normal order of things” (p. 684) to test poorly defined content targets, and it occurs whenever assessments are more concerned with measurement and technical considerations of the tests than anything else. Because of this, he claims that MDI’s effect on the curriculum is to “fragment it, narrow it, deflect it, trivialize it, and cause it to stagnate” (p. 684). Ultimately, what Bracey fears, and what Popham argues need not be the case, is that MDI will always aim low, as in minimum competency tests, and will further entrench education in instructional fragmentation.

Bracey’s argument rests on the notion that MDI has led to minimum competency tests that have narrowed the curriculum. He further asserts that MDI typically requires that assessments be “convenient,” which generally results in a multiple-choice format. He maintained that this causes test objectives – and subsequently by MDI’s design, instructional objectives – to become trivialized.

Yet, some researchers thought it worthwhile to investigate the concept of MDI using less convenient test formats. Shepard (1995) conducted a year-long project for in-service teachers to see how performance assessments in mathematics and reading might be used to “redirect instruction toward more challenging and appropriate learning

goals” (p. 40). However, Shepard advocated for a “bottom-up approach,” whereby teachers could experiment with new assessments and instruction, rather than having “high-stakes consequences” attached to new performance assessments imposed from above in order to “leverage change” (p. 38). Most notably, Shepard’s research underscored the importance of intensive professional development for teachers in order to effect assessment-driven reform in the classroom. Shepard stated that “it is hard to imagine how teachers could have gained such detailed project insights in a one-time in-service session or on their own, if external assessments were the only mechanism for instructional reform” (p. 42).

Despite such warnings, many states began including in their testing programs components that used alternative item types other than multiple-choice or short answer formats. In the Annual Survey of State Student Assessment Programs (Bond, Roeber, & Connealy, 1998), 21 states reported that as of the fall 1997 their assessments in mathematics included open-ended response items, and four states were using performance tasks in mathematics (with an additional four states using performance tasks in science). Two other states used examples of students’ work in their assessment measures. Nearly all states affirmed that instructional change was a motivator.

The “instructional change” referred to in the previous paragraph begs elaboration. The following section addresses the topic of mathematics instruction as it pertains to current trends in the United States because it is a key component of this study.

Mathematics Reform and Traditional Instruction

In the 1980s a litany of reports criticized the state of mathematics and science education. The National Commission on Excellence in Education (1983) called the United States “a nation at risk” due to an educational system that was producing high school graduates that were neither prepared for work nor for college. The report warned of the “rising tide of mediocrity” due to minimum standards required for graduation from high school and entrance to college. They specifically disparaged widespread minimum competency examinations because of their pernicious effect of dragging curricula standards down, as minimum competencies became de facto maximum objectives. Their broad recommendations for reform included a call for mathematics teaching to foster an understanding of concepts and enable students to apply mathematics in realistic contexts. The Mathematical Sciences Education Board of the National Research Council (1989) further chastised the state of mathematics education, blaming the traditional curricular emphasis on mechanics for the negative attitude toward mathematics so pervasive in our society and for the misconceptions among the public about mathematics’ utility in the world.

The reform efforts in mathematics were most vigorously heralded by documents of the National Council of Teachers of Mathematics (NCTM). Two of the original documents, known as the *Standards*, focused on curriculum, instruction, and teaching practices in mathematics (NCTM, 1989; 1991). In 1995 NCTM produced a third *Standards* document devoted to reform in assessment practices (NCTM, 1995). NCTM expanded on and consolidated its vision in its more recently published *Principles and Standards for School Mathematics (Principles and Standards)* (NCTM, 2000). The

Standards within *Principles and Standards* elaborate on specific content and processes on which mathematics instruction should focus for students in prekindergarten through grade 12. The Principles, within *Principles and Standards*, on the other hand, describe the “basic precepts that are fundamental to a high-quality mathematics education” (p. 6). Although, to some degree, all of the NCTM Principles are applicable to this study, those addressing learning and teaching are most relevant to the present discussion of reform instruction.

The Learning Principle emphasizes the importance of learning with understanding. It further makes the point that mathematical proficiency rests on a tripod of conceptual understanding, factual knowledge, and procedural facility, all of which are intertwined. The Teaching Principle emphasizes the importance of teachers knowing deeply the mathematics they teach, so that they can be flexible in organizing their instruction. A teacher’s task is to establish a supportive learning environment, while utilizing productive discourse in the course of implementing “worthwhile mathematical tasks.” Instruction should draw out what students already know and then build upon that. These principles are grounded in educational research and research in cognitive science (NCTM, 2000).

Although the term “constructivism” is notably absent from NCTM’s *Principles and Standards* (2000) document, constructivism is nevertheless the underlying theory of knowledge that is implied in NCTM’s principles. The main tenets of learning theory based on constructivism are that learners interact with their environment to construct meaning; that there is a structure to knowledge and that knowledge is accessed and employed differently in various kinds of learning. Thus, instruction must be designed to

foster this meaning-making, or “learning with understanding” as NCTM describes it. Hiebert and Carpenter (1992) explicated the idea of “learning with understanding” in the following manner: Knowledge is represented internally; internal representations are connected; and internal representations and their connections are influenced in some way by external representations. Therefore, if understanding is a dynamic process, and to understand something implies that “it is a part of an internal network” (p. 67), then external representations make a difference to the learners’ internal representations or understanding. This theory implies that learners process new information within the framework of what they already know, which consequently has implications for teaching for understanding.

In the 1990s, stimulated by the challenge of the 1989 *Standards* and drawing on an interpretation of learning influenced by constructivist theory, educators offered a new framework for teaching that was termed “reform instruction.” In mathematics education, reform instruction is often characterized by an inquiry-based learning approach, whereby students and teacher interact in a setting that emphasizes exploration and problem solving, conjecture, reasoning, and justification (Gregg, 1995; Brooks & Brooks, 1993). Rather than acting as omniscient authority figures offering primarily direct or summative statements in conjunction with demonstration, teachers use questions to focus instruction, to catalyze exploration, and to frame discussions around concepts in order to guide students toward mathematical understanding. Key characteristics of reform instruction are opportunities for both student-to-teacher and student-to-student interaction, as well as reflection by learners (Cobb, Wood, & Yackel, 1990; Brooks & Brooks, 1993).

This reform instruction encompasses a pedagogical approach that stands in stark contrast to the tradition of direct demonstration instruction so prolific in American classrooms. This “traditional instruction,” which is supported by behaviorist learning theory, has characterized the typical American classroom for over 100 years. Gregg (1995) points to Cuban’s comprehensive study in which he considered thousands of classrooms of various subjects and grade levels from 1890 to 1980. He described the tradition of instruction that endured throughout this time as “teacher-centered.” In a mathematics classroom, this type of instruction generally follows a familiar routine: answers to homework are checked, teacher presents new material using examples that are meant to be followed in a prescriptive manner, and time is given for individual student practice. Often, there is only minimal discussion or interaction, and the emphasis is on rote learning of prescribed procedures, rather than on concepts (Gregg, 1995).

Although NCTM has vigorously lobbied for reform in mathematics instruction since the late 1980s, and has continued to produce documents and resource materials to support teachers in their efforts, traditional teaching practices have persisted (Firestone, Winter, & Fitz, 2000; Spillane & Zeuli, 1999). The reasons cited for such persistence are many. Some researchers point to teacher beliefs about the nature of mathematics as either a conscious or subconscious explanation for their resistance to reform practices (Battista, 1994; Reys, Reys, Barnes, Beem & Papick, 1998). Others maintain that deficits in subject matter knowledge prevent teachers from successfully implementing reform (Ball, 1996; Guskey, 1994). Other researchers have found that without adequate participation in the process for implementing reform (Bay, 2000), or adequate

preparation to use reform curricula (Schoen, Finn, Griffin, & Fi, 2001), teachers are either unwilling or unable to change their teaching practices. Still others argue that the pressures of producing student success on enduring traditional standardized tests have “won out” as far as compelling teachers to choose teacher-controlled traditional methods rather than the “murky waters” of reform strategies (Cohen & Ball, 1990; Wilson, 1990).

To reiterate the relevant issues discussed above, many states have been using assessment programs not only for purposes of accountability, but as a vehicle to shape curriculum and bring about instructional change. The changes toward which they were aiming are outlined in state frameworks documents. In many cases, these documents are modeled after the NCTM *Principles and Standards* (2000), which express a vision of mathematics learning and teaching known as “reform” mathematics.

Assessments in Arbor County School District

The research reported in this study was conducted in the Arbor school district, which encompasses an entire county. At the time the research was conducted, the state in which the Arbor District is located used a performance assessment (PA) at the end of grades three, five, and eight as part of its state-wide testing program, in addition to functional tests of basic skills that were graduation requirements for students. The performance assessment program had high stakes for individual schools, whereby performance was linked to State intervention or even reconstitution. Thus, both the performance assessment and the functional test were high-stakes assessments. Both

contained a criterion-referenced mathematics component, which was the focus of the present study.

Eighth-grade students in this state had been required to complete both of these tests until 2002, when the performance assessment program was discontinued. (The PA was terminated due, at least in part, to new federal-funding guidelines requiring that state-mandated tests provide data on individual students.) All eighth graders were required to complete the PA at the end of the academic year, and any students who had not already passed the functional mathematics test (FMT), which in some schools was most eighth graders, were also required to take that examination. In the Arbor District, many of the students who had already passed the FMT were enrolled in algebra, a course for which there was a districtwide final examination. Students who passed the algebra final examination were awarded one credit of high school mathematics.

One of the stated purposes of the performance assessment was to improve instruction and curriculum for all of the State's students. This study sought to determine if in fact there is a link between instructional practices in the classroom and achievement on the statewide performance assessment and whether that link is the same for eighth-grade students enrolled in different courses.

Rationale

The FMT and PA tests reflect differing perspectives of knowledge and learning. The FMT is a traditional minimum competency test of basic mathematics skills; it is multiple-choice in format and assesses computational skills on whole numbers, fractions, decimals, and percents. This content is generally expected to be known by

students by the time they complete a sixth- and/or seventh-grade mathematics curriculum. In contrast, the mathematics portion of the PA is an “authentic” assessment, designed to assess the State’s specified learning outcomes. These outcomes, which are closely aligned with NCTM *Standards* (1989, 1991), were established by groups of educators and curriculum specialists and represent what “students should know and be able to do... as a result of their educational experiences” (MSPAP document, 1998, p. 2). The State Learning Outcomes include both content knowledge strands and processes, similar to those contained in the NCTM *Standards* (2000). These “processes” in which students are expected to gain facility are problem solving, communication, reasoning, and connections. The State’s board of education explicitly wrote that the PA assessed higher-order skills, “such as supporting an answer with information; predicting an outcome and comparing results to the prediction; and comparing and contrasting information” (MSPAP, 1998). The PA included items that integrated content knowledge strands and required the various processes to be demonstrated.

Minimum competency tests such as FMT became very popular in the 1970s and 1980s. Researchers have since documented serious negative effects, such as narrowing the curriculum, affecting student placement (which has a disproportionately negative impact on minorities and disadvantaged pupils), and constraining pedagogy and the instructional decisions of teachers (Noble & Smith, 1994). In teaching to tests like FMT, teachers are less likely to use practices such as cooperative learning, calculator exercises, and projects (Romberg, Wilson, Khaketla, & Chavarria, 1992) and are more likely to use instructional materials that resemble the standardized test (Mathison,

1991). Moreover, these negative effects have been found to be more pronounced for lower-scoring pupils (Linn, 1993; Graue & Smith, 1996; Robinson, 1996).

Since the State had defined an 80% standard as the satisfactory ninth-grade pass rate for FMT, it is reasonable to assume that many eighth-grade teachers in the Arbor school district felt pressure to ensure high pass rates among their students on the FMT. According to the research cited above, a consequence of this might have been that teachers of students who had not yet passed the FMT were using skill-focused teaching practices to routinize recall of items aligned with a traditional basic-skills test format.

The PA, on the other hand, had a stated purpose of moving the instruction and curriculum for all of the State's students toward specified standards and learning outcomes. Indeed, the State Board of Education (MSPAP, 1998) maintained in their PA document that they expected teachers to improve the performance of their students and explicitly argued that "in this sense, 'teaching to' the test is good instruction" (p. 13). There is little doubt that the creation of the PA was an attempt to bring the teaching of mathematics in this state more in line with the vision of curriculum, instruction, and assessment communicated in the NCTM documents. Indeed, their *Standards* documents maintain that assessment should evaluate students' "ability to explore, conjecture, and reason logically, as well as the ability to use a variety of mathematical methods effectively to solve nonroutine problems" (NCTM, 1989, p. 5). Thus, it would seem that if schools were striving to meet the State standards for performance on the PA, there would be a need for the kind of instruction endorsed by NCTM, which is commonly referred to as "reform" instruction.

Thus, the Arbor District is in a state that requires eighth-grade students to complete a performance assessment and, for many of these same students, a minimum competency basic skills test during the same academic year. State documents that list standards and learning outcomes delineate content and processes to be taught in a format akin to NCTM's *Standards* (1989, 1991, 2000) documents. Those State documents essentially form the guidelines for an intended curriculum. However, many aspects of a school program come together to influence what *actually* gets taught in a mathematics classroom – that is, the implemented curriculum. Via the PA, State assessments were then expected to measure the degree to which students in mathematics classrooms had learned the content outcomes specified by the State. Via the FMT, State assessments measure students' facility with recall of skill-based mathematics procedures and definitions.

Many eighth-grade students in Arbor District were required to complete both the FMT and the PA in the same academic year. Still other students completed both the PA and a districtwide Algebra examination. Although high-stakes for individuals did not accompany the PA, it is not unreasonable to associate the concept of MDI with the State's performance assessment program because of the high-stakes implication of student performance for schools and school administrators. Yet, surveys of both middle- and high-school teachers have shown that when faced with a standardized test like the FMT, teachers tend to focus on basic elementary content to the exclusion of achieving deeper understanding of key mathematical concepts (Romberg, et al, 1992; Stake, 1995). Some might argue that teachers were receiving mixed messages regarding what and how to teach their eighth-grade classes. Thus, Arbor District provided an ideal

setting in which to investigate the classroom instruction of students facing various and, in some sense, “competing” assessments.

Instruction and Course-Level Assignment as Variables

The mathematics specialist in Arbor District stated that although District documents stress a traditional lesson structure throughout each of the content areas, the District Mathematics Office tries to promote reform strategies within that structure. District-supported professional development was often aimed at helping teachers implement methodologies endorsed by NCTM and adapting their teaching strategies and curricular focus to align with the state’s learning outcomes. Individuals in the Mathematics Office believed that such strategies were most effective in producing higher achievement on the State PA.

However, Arbor District does not use curriculum materials, guides, or resources that are specifically *Standards*-based. In the absence of a comprehensive reform curriculum and/or professional development focused on the rationale behind reform content and strategies, teachers often resort to methods of teaching with which they are most familiar, that is, a more traditional, directed style of teaching. Reys, Reys, Lapan, Holliday, and Wasman (2002) noted that “the implemented curriculum often closely mirrors the content and pedagogical approach presented in textbooks” (p. 75). Thus, if teachers depend on the tone of textbooks to guide their own in-class activities, textbooks may have a larger impact on the mathematics that students learn than might be presumed by mathematics leaders in Arbor District.

Not unlike most school districts, the Arbor school district tracks students according to academic ability. That is, by the sixth grade, students are placed in a mathematics course based on perceived ability. Generally, teacher recommendations, along with prior achievement on a standardized test of basic skills, determine student placement. By the time students reach eighth grade, each student is put into one of either three or four different mathematics courses. In each school, those courses offer different levels of mathematics content and rigor. Each course utilizes a distinct textbook that has specific content emphasis and implied instructional strategies. The District practice of tracking has the intention of helping students of different abilities achieve at their optimum level; yet for many reasons, tracking is potentially problematic.

As soon as students are relegated to different classes based on ability, there is cause for concern regarding whether or not all students have the same opportunity to learn the skills and concepts expected by the State learning outcomes. Darling-Hammond (2000) maintains that unequal educational attainment stems from unequal opportunity to learn. She further argues that the most important variable in opportunity to learn is student access to high-quality teaching. In a perfect world, all teachers would be of the highest caliber, so that teachers who are fully knowledgeable in their subject area would expose students in all courses to the best possible teaching practices. However, this is not the case. Inevitably, some teachers do not have the necessary background in their subject matter, or have less experience, or lesser credentials. Thus, one would hope that a school district that separates students by ability level does not

likewise “track” teachers according to their credentials or abilities and place the weaker teachers in the classrooms with the lowest-achieving students.

A host of research has put forth evidence that there is, in fact, differential teaching practices among classes of students of different ability-levels. Besides disparities among quality of teachers, some researchers suggest that a more authoritarian-style of teaching occurs in classes of low-achievers, or even in whole school systems of economically disadvantaged children (Gregg, 1995; Haberman, 1991). An often cited cause for this practice is teacher beliefs about the innate abilities of tracked students, which translate into teacher expectations and eventually teacher actions in the classroom. Gregg found that when a teacher “limited the students’ opportunities to participate,” she could “limit the opportunities for a breakdown in control” (p. 456).

Another reason that instruction may be different can be gleaned from Gall’s (1984) research on teachers’ questioning types. Gall noted that teachers’ questions could be characterized within two categories: recall of fact and those requiring independent thinking. She concluded that “emphasis on fact questions is more effective for promoting young disadvantaged children’s achievement, which primarily involves mastery of basic skills ... and emphasis on higher cognitive questions is more effective for students of average and high ability, especially as they enter high school, where more independent thinking is required” (p. 41). Although Gall suggests that both types of questions be used for lower-achieving students in order to “stimulate development of their thinking skills” (p. 41), if lower track classes emphasize basic skills content over

other content, their teachers, even if only subconsciously, might be asking only lower level (fact) questions, as argued by Gall.

Interviews and surveys regarding teachers' practices in different academic tracks have, for the most part, shown similar patterns. Firestone, Mayrowetz, and Fairman (1998) observed and interviewed teachers in Maine and Maryland, two states implementing performance assessments, and found that teachers admitted to teaching different skills and concepts to students in different tracks because of their belief that students in lower tracks would not be able "to handle more open-ended problem-solving and inductive reasoning until they have first mastered the basic computational skills they [would] need to employ" (p. 107). In a broad survey of public and private school teachers, higher-ability students were given more problems that had several solution methods, a teaching strategy recommended by reform mathematics. Also, the lower the ability level of the students, the more teachers reported using routine exercises. When assigning homework, higher-ability students were more likely than lower-ability students to be given projects or problems with no obvious methods of solution and problems set in unfamiliar contexts, two other recommended strategies of the reform movement (U.S. Department of Education, 1999).

Hence, with the issue of academic tracking comes another layer in the investigation of how instruction is influenced by competing State assessments. Two philosophically different assessments converge on eighth-graders in Arbor school district. These students have been placed into different courses by academic ability, and therefore are often in a class made up of all students who face either one or both State assessments. Again, if MDI is operating, it begs the question of how instruction occurs

differently in distinct course levels, since students at the various levels have different assessment content and processes to confront at the year's end.

Instructional Approach and Achievement

At the time of this research project, the State PA had been the most recently added component of the State's assessment program, and indeed was anticipated to help shape improvement in instruction. Instructional expectations, as outlined in State documents, were aligned with reform instruction. Therefore, to whatever degree reform instruction was occurring in different course levels, it is worthwhile to examine whether certain patterns of instruction were correlated with higher achievement on the state PA. Whereas research has shown that reform-oriented instruction does not weaken, and in fact often raises students' achievement levels on traditional standardized tests (Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Cobb, Wood, Yackel, Nicholls, Wheatley, Trigatti, & Perlwitz, 1991; Mayer, 1998; Simon & Schifter, 1993), there is less research on the link between reform instructional models and mandated high-stakes performance assessments.

Additionally, since students are tracked in mathematics in Arbor school district, if it is the case that particular instructional strategies are associated with higher achievement, one needs to consider whether those teaching practices are consistent across the course levels, or at least not associated only with particular courses. Students in all classes must have an equitable opportunity to learn the content expected on the State assessments.

Research Questions

Presume that reform instruction is the desired effect of the PA. Yet the basic-skills FMT remains a hurdle for many of the state's eighth graders. Furthermore, eighth-grade algebra teachers feel that their course's curriculum is so dense, that the PA content or instruction is competing for time in the classroom. Thus, this research project sought answers to the following main questions:

- i. What factors related to teacher, classroom or student characteristics are associated with the amount of reform instruction that teachers perceive they implement?
- ii. Does the amount of reported reform instruction differ for eighth graders in different academic courses? In what ways do reported instructional practices differ for eighth graders in different academic courses?
- iii. Is there a relationship between the reported level of reform instruction and students' performance on the PA?

Classroom Observations

This study also included an observational component that characterized the instructional strategies of three eighth-grade teachers considered "exemplary" by principals or mathematics content leaders. The intent of this qualitative section of the research was twofold: to characterize the teaching strategies employed by these teachers and to examine the extent to which the students in these classes had the opportunity to engage in activities that are consonant with the expectations of the State PA exam. More specifically, three research questions will be explicitly addressed following a description of the observations:

- i. What are the characteristics of instruction in the classrooms of eighth-grade mathematics teachers identified as “exemplary” traditional and reform teachers?
- ii. Are there any comparisons or contrasts in the high-track classrooms with respect to teaching for conceptual understanding, particularly for two levels with the same teacher?
- iii. Do students in the classrooms observed appear to have the opportunity to learn the skills assessed by the mathematics component of the PA?

Although the observations can not be used to statistically validate the other measurements used in this study, they do offer insight into the complex nature of teacher practices and classroom interaction. Furthermore, since the same teachers were observed teaching two different levels of mathematics courses, their practices were evaluated to determine if students in different tracks were exposed to the same kinds of instruction.

Key Terms

High-stakes tests: Tests whose individual student scores are linked to student-level decisions, such as promotion, awarding of credit, or awarding of degree and/or whose aggregated scores are linked to school-level rewards or sanctions.

Instructional model: The particular pattern of teaching strategies employed by a teacher, on a continuum of traditional versus reform instructional practices.

Opportunity to learn: “The instructional opportunities and access to resources that would enable students to develop the complex thinking and problem-solving skills

that are the targets of the new assessments” (Herman, Klein, Heath, & Wakai, 1994, p. 3).

Reform instruction/reform strategies: Teaching strategies that have an inquiry-based approach which “emphasizes exploration, conjecturing, proving, and problem solving on the part of the students” (Gregg, 1995, p. 443).

Standards-based curricula: “Curricula developed to reflect the content and instructional philosophy outlined in the three [NCTM] Standards documents” (Reys, et al., 1998, p. 42).

Standards-based Pedagogy/Assessment/Approaches: Teaching practices or assessments that are implemented according to the vision of the NCTM as described in their *Standards* (1989, 1991, 1995) documents and in the *Principles and Standards* (2000).

Tracking: The practice of separating students into different courses based on a measured level of achievement or perceived performance in school.

Traditional instruction/traditional strategies: The characterization of instruction that is teacher-centered, whereby students are expected to remember and reproduce facts and procedures demonstrated by the teacher, with limited on-task teacher-to-student or student-to-student interaction beyond stating answers to exercises or recall of facts.

Worthwhile mathematical task: A mathematical activity that has the potential to “call for problem formulation, problem solving and reasoning; promote communication about mathematics; stimulate students to make connections and develop a coherent

framework for mathematical ideas; engage students' intellect; [and] develop students' mathematical understandings and skills" (NCTM, 1991, p. 25).

Significance and Purpose

Together, the quantitative and qualitative components of this study add to the growing body of research examining linkages between instruction and assessment for students in different mathematics tracks. If one purpose of the PA is to change instruction, then clearly it is important to know to what degree the intended instructional model is actually implemented in classrooms.

Additionally, if mathematics educators continue to press for both reform instructional strategies and authentic-type assessments, further evidence is needed to establish whether reform instruction is related to better student achievement on performance assessments. Perhaps most importantly, it is incumbent upon educators to know to what degree all students have the same "opportunity to learn" the skills and concepts assessed by the PA, and to determine the reasons for, and to try to alleviate, any inequities. As one step in such an investigation, this study proposes to examine 1) the degree to which reform instruction is reported as being implemented in different course levels of eighth-grade mathematics and 2) the relationship between reported levels of reform and student performance on the PA.

Limitations

The scope of this study had several limitations. First, this study took place in only one district in one state implementing a State-mandated assessment program. Thus,

the largest possible sample was restricted to the relatively small number of eighth-grade teachers in the district, which represents only a subset of the population of teachers in the State.

Additionally, the study sought to report on classroom teacher practices. However, the measurement instrument used to gather information about teaching practices employed in Arbor District was a teacher survey. Any instrument itself can be limiting in that information about only those teaching practices specifically listed on the questionnaire could be garnered. Furthermore, the information collected was self-reported. These data are actually a measure of teachers' *perceptions* of their instruction, rather than of actual practice. This must be taken into account in any interpretation of the data.

Other limitations were unexpected at the outset of the study. Originally, I expected to obtain actual scores on the FMT, but later learned that such scores were unavailable. Instead, the only accessible data was a pass/fail notation for each student as recorded in the district's database. Also unavailable were students' grades in prior mathematics courses, which might have been a more appropriate covariate in several analyses.

Assumptions

This study assumes that teachers' practices can be consistently documented through self-reported data. It is assumed that teachers will answer the survey questions as honestly as possible, while recognizing that different teachers could interpret survey questions differently. It is additionally assumed that those reported data on instructional

practices could be quantitatively summarized to represent a degree of reform-practices implementation on a scale from 0% to 100% of class time that those reform instructional practices are employed.

In further helping me to identify teachers for the observational component, I assumed that the specialist in the Office of Mathematics had the same understanding of notions of reform mathematics as offered by NCTM and as interpreted by me.

CHAPTER 2: REVIEW OF THE LITERATURE

The focus of this study lies at the intersection of several major issues in mathematics education, namely, assessment practices, equity, instruction and achievement. After a general discussion of high-stakes testing in the United States, the review outlined in this chapter centers on the relationship between assessment and instruction, examines equity issues surrounding reform mathematics and assessment, including tracking, and finally addresses the relationship between reform instruction and achievement.

High Stakes Testing

Since the 1970s, when educational accountability laws first emerged, high-stakes testing has been a fact of life in public schooling and the subject of on-going debate. During the 1970s, the prevalent theories of teaching and learning were based on behaviorist psychology. The belief was that by mandating certain assessments, teachers would respond to certain rewards or sanctions resulting from their students' test scores. In terms of the learner, students would be required to repeat material until they had reached a level of mastery. In this way, a foundation of basic skills was established that would precede higher-level instruction (Noble & Smith, 1994).

Initial accountability requirements based on minimum competency tests of basic skills were used to ascertain whether or not a student had learned sufficient content as a baseline for awarding of a high school diploma (Bond, Roeber, & Connealy, 1998). It soon became evident that "minimum" competency became the end goal for many students. In fact, although passing rates and scores on such tests eventually increased,

this gain was accompanied by a decrease in expectations encompassing higher-order thinking skills (Shepard, 1991). Additionally, Mehrens (1992) maintained that in some cases, test scores increased for reasons that ultimately invalidated the test measures themselves: Scores increased as a result of dubious test-preparation strategies and instruction that too closely matched testing content and format.

Noble and Smith (1994) cite numerous studies that point to the direct correlation between high-stakes tests and teachers' instructional decisions and materials. For example, Mathison (1991) found elementary school teachers' practices changing to reflect specific content and processes on the state's fourth-grade science assessment. Additionally, on a large-scale survey eighth-grade teachers revealed that they in fact do change their teaching to better reflect "the form and character of the [mandated] tests" (Romberg, Wilson, Khaketla, & Chavarria, 1992, p. 62).

Eventually, a preponderance of negative effects of standardized skills-based tests became evident. These included outcomes such as: narrowing the curriculum; encouraging out-dated teaching practices; siphoning lower-ability students out of mainstream classes; causing discouragement and higher drop-out rates among at-risk youth; and framing teaching in ways that so closely matched the format and content of the tests that when different tests of the same content in a different format were administered, students' scores were significantly lower (Shepard, 1991; Worthen, 1993). Although these effects were most closely associated with tests of basic skills, some educators have applied these same criticisms to any norm-referenced multiple-choice test (Miller & Legg, 1993).

Popham (1987) and Bracey (1987) summarily expressed the major points of the debate over measurement-driven instruction. Whereas Popham considered tests to be a cost-effective way to change instruction, Bracey countered that the academic cost would be limiting assessments to written forms that were “convenient” with “trivialized” objectives. Yet, with the publication of NCTM’s *Standards* documents (1989; 1991), this debate began to take on new meaning as educators began touting the far-reaching benefits that performance assessments might offer. Performance assessments were purported to test that which educators really wanted to know about: what their students could *do* with the skills they were learning. Performance assessments are meant to measure higher-order thinking and problem-solving skills and processes. Not all reviewers were proponents of performance assessments, however, as researchers’ warned of the drawbacks and possible misuses of performance assessments. Discussions of the psychometric properties of performance assessments questioned their generalizability (Mehrens, 1992) because of issues of equity, validity, and the genuine difficulty of trying to teach “students to use [higher-order thinking] skills that we are not certain how to teach and have limited knowledge of how to test” (Miller & Legg, 1993, p. 10).

But the idea of “measurement driven instruction” was firm in the minds of state policy makers, as evidenced by the fact that “improving instruction” was the most common reason given for state administered assessment programs (Bond, Roeber, & Connealy, 1998). Proponents of performance assessments argue that if the proposed linkage between the goals of instruction and performance assessment could be realized, then the once considered negative consequence of “teaching to the test” could be seen

as a beneficial tool for reforming curriculum and instruction (Linn, 1993). In fact, several states' rationales for their mandated performance assessments imply the expectation that the tests would help broaden, rather than narrow, the curriculum.

During the 1990s, many states instituted some kind of alternative assessment component within the state assessment programs. "Alternative" assessment refers to the use of a non-multiple choice format, such as extended response, performance task, short answer, or multiple choice with explanation (Bond, et al., 1998). Linn (1993) noted that these types of assessment programs would serve two main purposes: "to document the need for change...[and to act] as critical agents of reform" (p. 1). The first reason refers to setting standards for schools to achieve and holding accountable those schools that do not achieve as expected. Linn questioned the rationale that assessments would effect change as he called for critical analysis by the educational research community. The promises of performance assessment include increased student and teacher motivation, instructional innovation and constructivist orientation, and a genuine assessment of higher-order thinking (Linn, 1993, Noble & Smith, 1994). Linn argued, though, that it is incumbent upon the claim makers to demonstrate these purported salutary effects.

Performance Assessments as an Instrument for Educational Change

Several studies have been conducted to assess changes that have taken place in mathematics classrooms as a result of the administration of state performance assessments. Most of these studies have used teacher surveys as their main source of data, while some studies have used classroom observations and teacher interviews.

In comparing the effects of state performance assessments on the mathematics instruction in classrooms in Maine and Maryland, Firestone, Mayrowetz, and Fairman (1998) concluded that the relatively high-stakes pressure in Maryland resulted in somewhat more changes, however superficial, in that state. Overall, changes in instructional practices were not found. Instead, the authors felt that the Maryland State Performance Assessment Program (MSPAP) had only surface-level effects on teachers and, with regard to content, they noted a narrowing of the curriculum. Even when MSPAP-type activities were used, they were approached only perfunctorily, “accomodat[ing] deep-seated approaches to mathematics teaching” (p. 111). They noted many obstacles that stood in the way of influencing teachers’ strategies, among them teacher content knowledge and curriculum. Based on their classroom observations and interviews with teachers from four schools in Maryland, the authors found that the teachers were not concerned enough with sanctions to make real changes. They cautioned that their conclusion may not apply to urban areas where pressure to succeed is highest since sanctions in these districts were a real threat. However, they conjectured that in those areas there may be even more pronounced negative effects of testing.

A follow-up study compared jurisdictions in two countries on the question of whether policy, by way of assessment, could alter teaching practice (Firestone, Winter, & Fitz, 2000). The authors accessed data collected in the prior study in Maine and Maryland, and then conducted similar fieldwork in England and Wales, where there was a new national assessment. The conclusion was discouraging in nearly all cases. Except in Maryland, where as stated earlier there was a small, superficial change in content, there were essentially no changes in teaching methods. The authors conjectured that the

assessments did not provide enough leverage to stimulate what was needed to bring about real change in classroom instruction. That is, the authors felt that ultimately there was insufficient professional development opportunities to address the “pattern of teaching [that] appears to be deeply embedded in teachers’ ways of thinking about their work” (p. 18) and this hampers teachers’ efforts to alter their approaches. A drawback of their study, though, was the very limited sample size in each jurisdiction. For example, only 11 teachers from across four schools in Maryland were interviewed.

An earlier study accessed a broader sample of Maryland eighth-grade teachers, referencing 148 teachers from across the state, in order to study the perceived effects that the MSPAP had on instruction (Koretz, Mitchell, Barron, & Keith, 1996). Their reported negative effects included a perceived pressure to teach to the test and a lack of integration between the MSPAP and their curriculum. Furthermore, more than two-thirds of the teachers reported that the test caused some teachers to de-emphasize or altogether neglect untested material. On the other hand, 75% of the responding teachers reported placing greater emphasis on many content areas that are commensurate with new curricular goals, such as data analysis and problem solving. The surveyed teachers also reported many positive effects that MSPAP had on their instructional practices. They placed greater emphasis on writing, problem solving, thinking skills, real-life applications and hands-on activities. Teachers reported that they were more likely to use cooperative learning and to focus on process skills. The authors of the study noted a discouraging irony in their responses, however. When asked about general and specific test-preparation practices, teachers and principals reported placing greater emphasis on higher order thinking, using practice tests and setting higher expectations. Yet when

asked about the reasons for MSPAP gains, the least-cited response was broad improvements in knowledge and skills, with the most credit being given to direct test-preparation practices that could inflate scores. Whereas the report by Koretz, et. al. (1996) was more positive in terms of MSPAP's effects on teacher practices, the previously discussed study by Firestone, et. al. (1998) concluded that no real changes had occurred in Maryland.

Stecher, Barron, Kaganoff, and Goodwin (1998) conducted a similar study in which they surveyed teachers across Kentucky, reporting on the effects of the standards-based assessment in that state. With reference to a previous cycle of the state assessment, in both high-gain and low-gain schools, it was found that even though “traditional teaching approaches continued to be used on a regular basis, the greatest change in practice was in the direction of standards-based approaches” (p. 26). Some findings are particularly interesting in light of the present study. With specific regard to eighth-grade mathematics, the researchers found that more teachers in high-gain schools than in low-gain schools increased their use of extended investigations and non-routine problems when assessing student performance. These high-gain schools avoided the use of multiple-choice tests altogether. A greater percentage of high-gain schools than low-gain schools reported covering algebraic concepts regularly. Although more high-gain schools seemed to quiz or test more frequently, they also were more likely to count performance tasks in grading. In high-gain schools there was a greater percentage of teachers reporting that school staff, district resource staff, and NCTM *Standards* documents had a strong influence on their content and practices. Twice as many teachers in high-gain schools reported increases in remedial and enrichment support for

their mathematics students. Although it is tempting to credit the reform-oriented approaches with the higher-gains made on the performance assessment, the authors did not control for other variables that could have had an impact on student achievement gains.

The National Center for Education Statistics reported results from a national survey of 4000 public- and private-school teachers in elementary and secondary schools (United States Department of Education, 1999). In light of the debate over instructional philosophies that took place in the 1990s, this survey sought to gather information about teacher practices. In general, teachers reported using many practices consistent with recommended *Standards*-based pedagogy. An interesting result is that for most of the “recommended” teaching practices, more teachers of low-ability students as compared to teachers of high-ability students reported implementing those practices. The exceptions were that higher-ability students were given more problems that had several solution methods, and the lower the perceived ability of the students, the more teachers reported using routine exercises. This supports the notion that high-ability students are more likely than low- ability students to be exposed to higher-order thinking activities. When surveyed about homework assignments, teachers reported that the higher the perceived student-ability, the more likely those students were to be given projects or problems with no obvious method of solution or unfamiliar problem situations.

Except for the study by Firestone, et. al. (1998), the other studies cited above depended on teacher self-report data. A more in-depth observational study was conducted by Spillane and Zeuli (1999). In this study the authors used the TIMSS survey data to identify 25 mathematics teachers who reported using instructional

practices consistent with reform and who indicated familiarity with national or state standards. The authors were focused on “principled mathematics knowledge” (p. 4) as opposed to procedural knowledge, which they determined could be ciphered from teacher questioning and classroom discourse. Thus, they looked beyond the surface activities and materials used in the classroom. They found that in fact the “behavioral regularities of instruction” were more malleable than the “epistemological regularities” (p. 19). In other words, although in fact teachers made use of manipulatives and cooperative learning and encouraged interaction in the classroom, many of the teachers still conducted the class as “a quest for the right procedural answer”(p. 20). They determined that only 4 of the 25 teachers conducted their classrooms within the spirit of the mathematics reforms. Another 10 teachers presented tasks that had the potential to highlight conceptual knowledge, but classroom discourse fell short of drawing out the principled knowledge. The remaining 11 teachers used tasks and practices that were firmly grounded in procedural knowledge. Spillane and Zeuli concluded that teachers must learn and grapple with the concepts behind reform proposals, and that policy alone would not effect change in patterns of practice in classrooms.

In their studies of teachers and classrooms, Firestone, et al. (1998) and Spillane and Zeuli (1999) found what Cohen (1995) had summed up earlier. Considering the research thus far, Cohen disputed the assumption that policy instruments, such as state standards and assessments, could “drive” instruction and result in real systemic reform at all. He maintained that instruction is a by-product of teachers’ knowledge, their professional values and commitments, and their “social resources of practice” (p. 16).

He further argued that in the United States those elements are both weak and weakly related and governmental policies have done little to address the weaknesses in them.

Barriers to Reform Instruction

In explaining the many reasons why reform has not taken hold, specifically in mathematics instruction, Cohen (1995) pointed out that teachers “are much more literate than numerate” (p. 14); thus, while language arts teachers are able to make sense of reform recommendations and better utilize available innovative materials, this has not been the case in mathematics.

Battista (1994) argued that the problem in implementing *Standards*-based teaching stems from teacher beliefs about mathematics. That is, teachers maintain the belief that mathematics is a set of rules and procedures to be learned, rather than the idea that mathematics is a sense-making endeavor. Case studies looking at the extent to which reform curricula are or are not implemented have concurred with Battista’s conclusion about teacher beliefs. Rickard (1995) studied a middle school teacher who used a unit from a problem-oriented mathematics curriculum. He found that her beliefs and content knowledge interfered with her ability to implement the curriculum in the way it was intended by the authors. He found that she “maintained her focus on problem-solving as the application of rules and procedures” (p. 22). When looking into obstacles in the way of reforming mathematics instruction, Reys, et al. (1998) likewise found a distinction between teachers’ stated beliefs and their “beliefs-in-practice.” This distinction might explain why Jacobs and Morita (2002) found that American teachers were more “flexible” in what they considered good teaching practices, compared to

Japanese teachers, who were more unwavering in their expectation for a student-centered conceptual-development approach in a mathematics lesson. In comparison, the American teachers were “supportive of both nontraditional and traditional elementary school mathematics instruction” (p. 172), even for teachers who were familiar with reform documents. The authors of the study considered this to be indicative of the fact that the current theory of learning espoused in mathematics education has long preceded actual implementation of that theory in the classroom; hence, the often seen disconnect between teachers’ beliefs and their practice.

Aside from teacher beliefs however, both Reys, et al.’s and Rickard’s studies also pointed to weaknesses in content knowledge *in combination with* beliefs about mathematics, that resulted in ineffective use of *Standards*-based curricula or practices. Guskey (1994) argued that even when teachers had a positive attitude regarding performance assessments, their lack of knowledge about how to adequately prepare students for that type of test, as opposed to a traditional basic skills test, and how to manage within their perceived classroom time constraints prevented them from addressing the needs of the students.

Many other studies show the importance of teacher buy-in as an essential variable in the extent to which reform curricula are effectively and appropriately implemented in the classroom (Bay, J. M., 2000; Reys, Reys, Lapan, Holliday, & Wasman, 2003; Riordan & Noyce, 2001; Schoen, Finn, Griffin, & Fi, 2001; Stein, Grover, & Henningsen, 1996). Many of those studies highlighted the importance of teacher preparation in using a comprehensive reform curriculum prior to implementation in order to make the greatest impact on students’ achievement. This

finding was determined based on the outcome that growth in student achievement was positively correlated with teacher preparation (Schoen, et al., 2001) or length of time in the particular reform curriculum being used (Riordan & Noyce, 2001). Bay's (2000) case study of six teachers in two districts that were part of an effort funded by the National Science Foundation (NSF) to change over to *Standards*-based curricula pointed to the direct correspondence between the teachers' professional development and level of participation in the decision-making process and the level of implementation of *Standards*-based practices. As level of student inquiry, classroom discourse, and focus on concepts over computation was considerably reduced, the less the teachers participated in the out-of-classroom opportunities for collaboration.

Teachers, themselves, cite another reason for the lack of reform-practice implementation in some classrooms. In Firestone, et al. (2000) teachers reported having different expectations for students of different ability. These teachers felt that lower-ability students lacked the sophistication necessary to learn mathematics using many of the practices recommended by the reform movement. This reason brings another major educational issue – the issue of equity and tracking – into the picture.

Equity Issues in Assessment

There are many causes for concern that the achievement gap between the majority group and under-served minorities could stay the same or even grow with the advent of performance-based assessments. Some educators believe the variable that could have the most impact on equity is “opportunity to learn.” As defined by Herman, et. al. (1994), opportunity to learn is “the instructional opportunities and access to

resources that would enable students to develop the complex thinking and problem-solving skills that are the targets of the new assessments” (p. 3). Linn (1993) argued that if past inequities in educational opportunities do not change, so that the achievement gap remained, “the resulting disparate impact on minority students will not only undermine the system but also demonstrate a failure to achieve the goal of providing better education for all students” (p. 8). Thus, in calling for research on the consequences of performance assessment programs, Linn specifically reminded researchers to consider the *unintended* consequences, particularly with regard to issues of equity and bias.

One equity issue is the fact that performance assessment requires students to be competent readers and writers, even when they are being assessed on mathematics. Koretz, et al. (1996) found that most teachers believed that the emphasis on writing made it difficult to fairly assess their students’ mathematical competence, which might call into question the validity of a performance assessment. However, the authors appropriately point out that if test developers believe that mathematical communication is an integral part of mathematical competence, then writing would not necessarily lessen the validity of the test.

A study by Herman, et al. (1994) investigated whether or not equal opportunity to learn material assessed by a California performance assessment existed across urban, rural, and suburban middle schools. Surveys, observations, and interviews with teachers and students revealed that although urban students in the sample were exposed to the same kinds of instructional practices, and thus “were not limited to a meager ‘drill and kill’ curriculum,” they did in fact have less access to recent texts, were less likely to

understand a “key concept in mathematical thinking” (p. 45), and felt less prepared in general for the assessment. A limitation of that study, however, was their small sample size due to the classroom as the unit of analysis. As a result, their finding of no significant differences in instructional practices could very well be due to the constraint on the power of their analyses. Another study by Herman and Golan (1993) looked at the differential effects of standardized testing on schools with increasing scores and schools with stable or decreasing scores. Findings showed that schools serving more disadvantaged students followed a narrowed curriculum and that they emphasized more testing and test preparation. Robinson (1996) referred to this as “the tyranny of low expectations” (p.389) that plague, in particular, urban schools. She pointed to National Assessment of Educational Progress (NAEP) data indicating that urban students completed less challenging mathematics courses and were less likely to pass through the algebra “gatekeeper” course. Robinson blamed both low expectations and the current basic-skills assessments for blocking students from “rich and challenging” mathematics curricula.

Besides inequities between schools, just as important is the equity issue raised in the context of tracking students *within* schools. The next section addresses the issue of tracking as it relates to student achievement and opportunity to learn.

Tracking and Opportunity to Learn

Tracking is defined as separating students into different courses based on a perceived or measured ability level. Generally, there are two focuses in the debate over

tracking: 1) its efficacy in terms of student achievement, and 2) its effect on issues of equity.

In his review of 29 studies conducted to determine the effect of tracking on student achievement, Slavin (1990) concluded that students in same-ability groups fared no better than students that were in mixed-ability groups. He determined this based on the fact that some studies showed positive effects, while others showed negative effects. Thus, when he essentially aggregated the data, tracking was reported to have no overall effect on achievement or inequality. Where significant differences in individual studies were found, Slavin argued that it was due to measurement error.

Other studies revealed another typical pattern of results that leads to the “no effect” conclusion. In some studies, while students of high ability benefited (though not always to a significant degree) from homogeneous grouping, the significant losses in achievement sustained by low-ability students in homogeneous settings offset any gains seen in the high-ability groups (Hoffer, 1992; Kerckhoff, 1986). For example, Peterson (1989) conducted a study to compare achievement among remedial students (as identified by Comprehensive Test of Basic Skills and Intelligence Quotient scores) that were placed in one of three programs for seventh-grade mathematics: 1) a homogeneous low-track class, 2) a homogeneous average track, or 3) an “accelerated class” of mixed-ability students. At the end of the year, all students were tested in computation, problem solving, and mathematical concepts. Those in the mixed-ability accelerated classes showed significantly more improvement in all three areas than their counterparts in either of the homogeneous classes. Perhaps an even more striking result was that when they re-tested the students after the courses, 19 of the 100 remedial students that were in

the advanced program tested into a higher ability category; none of the other 200 students in the low and middle tracks did so, and the gap dividing the three categories was wider. Thus, not only did mixed-ability grouping help remedial students learn more, but it also provided advancement opportunity, as opposed to merely perpetuating the status quo.

Linchevski and Kutscher (1998) also found that intermediate- and lower-ability students performed significantly worse when they were homogeneously grouped than when they were in mixed-ability settings. The lower-ability students in their study's heterogeneous classes seemed "accustomed to much higher demands and expectations" (p. 545). Higher-ability students in homogeneous groups fared better, though to a non-significant degree, than their counterparts in mixed-ability settings. Thus, the authors concluded, the larger gap that occurs in tracked systems is due to the loss in achievement of the lower-ability students, not from gain of the higher-ability students.

The achievement gap that exists between tracks is unfortunately often correlated with social class. In a study on reform and equity, Boaler (2002) came across a noteworthy incident. She studied students who had attended a middle school that had used tracking with a traditional curriculum. At the end of middle school, the correlation between social class and mathematics achievement was .43. After three years in a high school that used mixed-ability grouping with an open-ended, projects-based curriculum, the correlation between social class and educational attainment, after controlling for initial attainment, decreased to only .15. She also looked at another group of students who had the opposite schooling experience; they went from a mixed-ability middle school, where the correlation between social class and achievement was .19, to a

traditional curriculum in a high school that tracked by ability. The same correlation at the end of high school had increased to .30. Boaler's finding seems to lend support to the notion that tracking could be responsible for exacerbating the relationship between class and educational achievement; however, due to the apparent confounding with type of curriculum, it is impossible to make a definitive claim regarding tracking's unique influence.

Gamoran (1992) takes issue with Slavin's (1990) conclusion that the net effect of tracking is zero. Rather than attributing the net result to measurement error that was the cause of finding a mix of positive and negative results, Gamoran argues that "ability grouping has different effects depending on where and how it is implemented" (p. 13). That is, he contends, "ability grouping has no effects on achievement unless teachers use it to provide different instruction to different groups" (p. 13). Thus, tracking inevitably raises questions about equitable instruction across classrooms, the second major focus of tracking research.

Oakes (1986) has argued vehemently against tracking, citing differences not only in educational outcomes, but differences in instruction as well. Referring to the higher quality of instruction that higher track classes are exposed to, she explains that those classes are more desirable to teach, so that better teachers are rewarded with those classes. Such teachers are often more enthusiastic and spend more time preparing (Oakes, 1991). Darling-Hammond (2000) agrees that unequal educational attainment stems from opportunity to learn and goes further to say that the most important variable in opportunity to learn is student access to high quality teaching.

Examining the quality of instruction in different tracks of eighth and ninth grade classes, Gamoran (1995) found significantly different questioning patterns among honors, regular, and remedial classes. Differential student participation and discussion indicated that interaction between student and teachers also varied by track. Other studies on tracking and educational achievement also sought to find reasons for the negative results found for low-achievers in homogeneous settings. Peterson (1989) found that the low-track classes were characterized by lecture mode since the students asked few questions. Furthermore, a negative attitude about mathematics pervaded the classroom atmosphere of the low-track classes. By contrast, the mixed-ability advanced classes had more discussion and an abundance of student questions.

Why do so many teachers fail to teach higher-order thinking skills? Class level (or track) of the class being taught explained nearly all variation in the amount of higher-order thinking skills taught in a study by Raudenbush, Rowan, and Cheong (1993). Grade level explained variation further, but this is not unusual due to the hierarchical nature of the mathematics curriculum in the United States. That is, more advanced mathematics courses are taught in later grades, and it is in these mathematics courses that more high-level and abstract thinking is expected.

A reason cited for differential treatment of lower-ability classes, also mentioned in an earlier section of this review, is the differing expectations regarding what mathematics content teachers feel is manageable for students of varying ability levels (Firestone, et. al, 1998, 2000). In a case study, Gregg (1995) cited the teachers' and principal's life/career expectations for the lowest-track students as a reason for instituting an authoritarian style of teaching. In their interviews with teachers in Maine

and Maryland, Firestone, et al. (1998) found that teachers taught different skills and concepts to students of different ability levels because they believed that mathematics was hierarchical; students had to first master the basics of computation before being able to handle “more open-ended problem-solving and inductive reasoning” (p. 107). Mayer (1998) found that eighth-grade teachers of algebra were more likely than ninth-grade algebra teachers to use “NCTM-style instruction.” He suggested that this could be an artifact of teachers’ perceptions that eighth-grade algebra students have higher mathematical ability simply by virtue of their advanced status. Thus, students who never even make it to or through algebra are often cut off from the opportunity to engage in higher-order learning, since early access to algebra was found to have an impact on “the socialization of students into academic work over and above simple exposure to content knowledge” (Smith, 1996, p. 150).

Yet another factor that could exacerbate the circumstance of differential instruction across tracks could be the discrepancy in high-stakes assessments required by students in different courses. Indeed, the state in which the present study was conducted requires the passage of a minimum competency mathematics test, the FMT, in order for students to receive their high school diploma. There is considerable pressure for students to have already passed this test by the time they reach high school. Many eighth-grade students in this district, and particularly the ones in the lower tracks, still face that hurdle. Shepard and Dougherty (1991) conducted a study on the effects that such multiple-choice format standardized tests have on instruction. Extensive teacher surveys revealed that teachers tended to emphasize basic skills instruction due to pressure to increase test scores. Other studies have also shown that high-stakes tests

have the unintended consequences of including in the curriculum that which the test covers to the exclusion of other material. For students who have yet to pass the FMT, often concentrated in the lower-track classes, this would mean a curriculum focused on very narrow and shallow mathematics content. Furthermore, minimum competency tests have been found to limit placement opportunities for students who score low (Smith & Rottenberg, 1991). Therefore, the influence that minimum competency tests can exert on the curriculum, and particularly so for lower tracks, might act as a roadblock to more advanced mathematics for many students.

There is another side to the relationship of high-stakes testing and tracking and the subsequent effect on instruction. Where there exists a performance assessment for all tracks, teachers do not always believe those tests offer the right opportunities for students in lower tracks. Not all teachers of all students agree with the argument that performance assessments and compatible curricula provide greater intrinsic motivators for students to learn mathematics. In their comprehensive report about the perceived effects of a state performance assessment, Koretz, et al. (1996) found that “far more teachers (particularly eighth-grade mathematics teachers) reported that expectations had increased greatly for high-achieving students than for [lower-achieving] groups”(p. 26) since the implementation of that state’s performance assessment. Doyle (as cited in Crooks, 1988) suggested that one reason why teachers test lower cognitive skills is that the teachers believe that using higher level questioning would lead to anxiety and frustration for many students, and ultimate failure.

Having cited research that linked minimum competency tests with dropout rates, Linn (1993) then considered how performance tests might affect student motivation.

That is, perhaps low-achieving students might see the new standards as well beyond their abilities and become even further discouraged. In his review of the research of the effects of classroom evaluation on students, Crooks (1988) cited researchers who suspected a curvilinear relationship between higher standards and student effort; at some point, the weaker students become discouraged and withdraw from the learning process altogether.

Thus, teachers appear to have different expectations for different students and some might believe that the cognitive demands that performance assessments place on students might actually undermine, rather than enhance, students' motivation. Thus, for many reasons, the record has shown that teaching toward lower-level thinking skills is more prevalent among lower-track classes.

Effects of Reform Instruction on Achievement

Traditional teaching styles characterized as teacher-directed, replete with individual seatwork, and that are textbook- or worksheet-driven have been associated with a lower cognitive demand level. Reform approaches are generally understood to be more cognitively demanding and aimed at developing higher-order skills, such as "analysis, synthesis, complex problem solving and oral expression" (Worthen, 1993). Analogous to instructional varieties are assessment types. Traditional tests of multiple-choice or single-answer items are criticized for measuring too narrow or low-level of a thinking ability, whereas performance assessments are purported to tap higher-level thinking skills. An assumption that teachers make, whether consciously or not, when they choose to emphasize drill and practice on basic skills to the exclusion of more

challenging instructional tasks, is that students would not perform well on traditional tests if they are not taught in a traditional manner. Over the past decade more and more research in mathematics education has been aimed at clarifying the link between the kind of instruction (traditional or reform) to which students are exposed and their subsequent performance on different kinds of assessments (traditional or performance). Indeed, a component of the present study seeks to add to that body of research.

Studies set in various grade levels have sought to answer the question about how students perform on different types of assessments, after being engaged in different kinds of instruction. Several studies with primary grade students show that not only does reform teaching *not* hinder students' performance on traditional tests, but that in many instances, it leads to better performance, particularly on more conceptual items (Carpenter, et al., 1989; Cobb, et al., 1991; Simon & Schifter, 1993).

In a study focused on secondary mathematics, algebra growth as evaluated by traditional tests was found to be evidenced sooner in classes where teachers used more reform teaching practices. However, while reform teaching fostered faster growth with high-achievers, it did not do so for lower-achievers. Yet, in this study the reform approach did not lower any groups' performance (Mayer, 1998). Similarly, Wiley and Yoon (1995) found that curriculum and instruction which was aligned with a more reform approach, one that fostered higher-level thinking, correlated positively with student scores on the test of the California Learning Assessment System. This result also implied that students who came from classes which emphasized less higher-level thinking skills via more traditional instructional strategies did not perform as well on the assessment.

Several studies have looked at student achievement in curricula that were specifically created as *Standards*-based curriculum projects. For eighth-grade students who spent at least two years in one of two different *Standards*-based curriculum, achievement on the Missouri Assessment Program (MAP) exam equaled or exceeded achievement of students in comparable districts who used traditional curriculum materials (Reys, et al., 2003). Furthermore, all of the groups in the *Standards*-based curriculum scored significantly better than the comparison groups on the algebra portion of the MAP, providing evidence that contradicts the argument that reform-based curricula are weak on developing algebraic skills. The MAP test was developed to assess the mathematics in the state Framework, which is based on NCTM *Standards*, and this test addressed skills, concepts, and problem solving, including both open-ended and multiple-choice questions.

Another study also examined the effect of elementary and a middle-school *Standards*-based curricula on student achievement on the mathematics portion of the Massachusetts Comprehensive Assessment System (MCAS) (Riordan & Noyce, 2001). Similar to the MAP, the MCAS was based on the Massachusetts Curriculum Framework, which was designed to be aligned with NCTM *Standards*. MCAS included both open-response and multiple-choice items. Again, these authors compared students using the reform curricula to comparable students using traditional textbooks in other schools. They noted significantly better student achievement from the *Standards* curricula across students of different gender, race, and economic status, and with only a few exceptions, students in the *Standards*-based curricula outperformed the comparison groups in all four areas of mathematics on all three types of questions. (The exceptions

were that no significant differences were found between the traditional and *Standards*-based elementary curricula in two areas of mathematics, and the comparison group outperformed the early-implementing *Standards*-based middle school curriculum on short-answer items.) Furthermore, schools that had implemented the *Standards*-based curriculum for a longer period of time had sustained or increased gains when compared to baseline data.

Another research project investigated the efficacy of a *Standards*-based curriculum, but used a study-developed assessment to separate out kinds of algebra knowledge. Huntley, Rasmussen, Villarubi, Sangtong, and Fey (2000) assessed students who had used the Core-Plus Mathematics Project (CPMP) as well as students in a traditional curriculum, comparing students on various mathematical achievement outcomes in the algebra and function strand. The results were mixed: Students in the CPMP were able to better solve realistic, context-bound problems when graphing calculators were allowed; students in the traditional curriculum fared better than their counterparts in CPMP on non-contextualized, symbol-manipulation exercises when graphing calculators were prohibited. These results highlight the need to clarify what mathematics learning is most valued.

Some researchers have cautioned against using reform approaches with lower-achieving students, citing problems that reform instruction has posed for those students (Baxter, Woodward, & Olson, 2001). Some of the researchers even suggested that using such approaches could worsen the achievement gap and hence, promote more inequity (Lubienski, 2000). However, many other studies have found *Standards*-based curricula, or in some cases particular reform approaches, to be beneficial to low achievers.

The study by Riordan and Noyce (2001) discussed above found that even when they separated the students by prior level of achievement and compared students with their achievement-level counterparts, students using the *Standards*-based curricula fared significantly better than those who had used the traditional textbooks *across all levels of achievement*. In that study, gains could not be attributed to differences in teacher qualification or instructional practice, or even level of curricular implementation. That is, the only variable considered was the textbook series being used – *Standards*-based or traditional.

Ginsburg-Block and Fantuzzo (1998) studied the effect of two instructional approaches advocated in reform literature – problem solving and peer collaboration – on 104 low-achieving third and fourth graders. Each of the two approaches, individually, resulted in significantly better outcomes on a range of student measures, both mathematical and affective. Surprisingly, however, a treatment using both interventions did not provide any significant benefit beyond what either individual intervention offered. In a study with low-achieving third through eighth graders, Cardelle-Elawar (1995) noted that low-achievers benefited by deliberate metacognitive instruction in mathematics. Specifically, they outperformed control groups at each grade level on learning outcomes related to mathematics problem solving. Bottage (1999) compared remedial and average students' achievement after either contextualized-problem instruction or traditional word-problem instruction. For students of both ability-levels, those who received contextualized-problem instruction outperformed the comparison groups.

In comparing secondary students' achievement from two low-income areas in England, Boaler (as cited in Boaler, 2002) found that students using an "open-ended approach" curriculum compared to a traditional curriculum attained "not only a higher level of achievement but also more equal achievement" (p. 246). Boaler's subsequent research found that those teachers who were well-prepared for and dedicated to teaching with reform-based curricula were able to reduce the inequalities based on race and class in their respective schools. Her findings focused on specific teacher practices that led to success: helping students understand questions posed to them, teaching students how to learn through communication and justification, and discussing with students contextualized problems and how to recognize various interpretations of them (Boaler, 2002).

The QUASAR Project addressed the question of how a minority population, and often low-achieving population, would fare when instruction and assessment were linked in their goal of helping students to gain mathematical power. The QUASAR Project was designed to bring alternative, reform-minded curriculum, instruction and assessment to urban middle schools. Silver and Stein (1996) reported substantial gains in students' mathematical problem solving, reasoning and communication as measured by QUASAR's performance assessment, the QUASAR cognitive assessment instrument (QCAI), over a three year period. Learning gains were substantial in classrooms that fostered high-level thinking through the use of instructional material that consisted of "worthwhile tasks." Yet gains were minimal in classes where instructional methodology was more traditional and based on teaching procedural skills. Additionally, QUASAR students outperformed students from a similarly disadvantaged urban sample on all

mathematics areas assessed by NAEP, particularly for questions that were more conceptual in nature (Silver & Stein, 1996). These students even performed better than a national NAEP sample on problem-solving tasks or tasks assessing conceptual understanding.

Summary

The significance of the research presented in this chapter poses several notions related to the present study: In order to perform well on the state's performance assessment, eighth-graders might benefit from reform-based instructional practices if their teachers are prepared to and capable of implementing those practices appropriately. In this district that tracks students into different courses by ability level, even typically under-achieving students stand to gain academic ground under those circumstances.

However, many students in some classes are still under pressure by needing to pass the minimum competency FMT, and still other classes face a district-wide final examination addressing algebra. All of these end-of-year assessments could be creating a tug-of-war on the classes' curricular organization and the teachers' instructional plans. Consequently, there might be serious obstacles to some students' opportunity to learn. That is, pressure to cover content on other tests besides the state performance assessment could conceivably limit students' exposure to important mathematics topics not assessed by either the basic-skills FMT or the algebra final examinations. Additionally, perhaps due to test pressure or a myriad of other reasons discussed in this

section, teachers might be using different instructional approaches with students in different courses.

What occurs in classrooms is undeniably related to what mathematics students learn. The PA in this state attempts to measure how much mathematics and what mathematics students have learned. In light of the importance policy-makers and stakeholders have placed on the connection between what occurs in the classroom and student achievement, this study sought to add to the literature surrounding that issue. This study endeavored to determine what kind of instructional strategies – reform or traditional –teachers tended to use more, whether this differed in different courses, and whether there was a corresponding impact on student achievement on the PA.

CHAPTER 3: METHODS

One of the main goals of this study was to determine if there was a relationship between the instructional strategies that teachers employed and the achievement level attained by their students on a state mathematics performance assessment. The methods described in this chapter pertain to the quantitative study undertaken to explore that relationship, as well as to characterize in general the instructional practices being employed in various tracks of eighth-grade mathematics classrooms. This chapter initially provides a brief overview of the methods employed in this study, with an elaboration on the target population and sample, instruments and variables, and the statistical analysis presented in the remainder of the chapter.

This study accessed the eighth-grade mathematics teachers in one school district. The sample consisted of those teachers who returned a completed two-part questionnaire and for whom the school district provided class rosters for use in this study. There were 51 teachers from 18 different middle schools across the district in the analytic sample. Analyzed class-level student data came from one class from each of those 51 teachers. The description and analysis that follow address school, teacher, and class-level background characteristics.

Population and Study Site

The target population for this study was eighth-grade mathematics teachers and one of their classes in a countywide school district in a state that administered an eighth-grade mathematics performance assessment (PA) as part of its statewide testing program. Arbor County, which comprises the third largest school district in a state of

24 school districts, was accessible and willing to participate in the study. The district serves 12.5% of the state's total student enrollment and is extremely diverse in its population. There are areas that could be classified as suburban, rural, and urban. Some schools are nearly all White; others are predominantly Black. Some schools serve an upper-middle class population, whereas others serve low-income areas. Thus, given the diversity in the middle schools, this district was considered in some ways to be an ideal setting in which to compare reform instruction in schools that serve very different demographic populations. There are 26 middle schools in Arbor County, with approximately 91 teachers of eighth-grade mathematics.

Sample

The sample of teachers and classrooms for this research was obtained through a multi-step process established at the outset of this study as a condition for permission to conduct the study in this district. First, principals of all 26 middle schools were sent a letter (Appendix A) asking for their support and explaining the study. The letters merely requested access to the school, so that their eighth-grade mathematics teachers could be invited to participate in the study. In nearly all cases a follow-up call was necessary to obtain a response. Of the 26 principals, 20 either allowed their teachers to be approached, or simply deferred a response to the Mathematics Content Leader (MCL), who was the next point of contact in any case. Six principals did not agree for these cited reasons: Their school or teachers were already involved (or had just finished) with other research projects, or they felt it to be an unnecessary burden on teachers following a recent administrative change.

Once given school access, I conducted a telephone interview with each school's MCL to gain an understanding of the particular mathematics program at that school and to obtain the names and teaching schedules of all teachers who taught at least one eighth-grade mathematics course. One MCL did not return my calls, however that school was retained in the sample, as the principal in that school answered my questions directly.

Subsequently, questionnaires were sent to 67 teachers in 20 middle schools. The teacher survey was separated into two questionnaires. One questionnaire was sent in November, and the other was sent in March. For both mailings, procedures were undertaken in order to ensure the highest possible response rate. First, I had enlisted the support of the MCLs to alert teachers that they would be receiving a questionnaire and to persuade the teachers to complete the questionnaire in a thoughtful and honest manner. Some of the content leaders even said that they would make time during a meeting for their teachers to complete it. Two weeks prior to sending the questionnaires, teachers were sent a postcard that briefly introduced them to the study and researcher and alerted them to the up-coming questionnaire. The questionnaire itself contained a cover letter explaining the study and stating that it had been sanctioned by the district research office, but was strictly voluntary on the part of the teachers. A prepayment of two dollars was attached to each questionnaire, as a token of appreciation. Fowler (1993) cites several studies indicating that this approach has been successful in increasing response rates for many populations. I sent a reminder postcard two weeks after questionnaires had been sent, and I sent another copy of the questionnaire to non-respondents after four weeks.

A brief examination of the returned questionnaires revealed that some respondents completely omitted answers for the most important and challenging question concerning instructional practices. These teachers were sent back their survey with another cover letter asking them to attempt to answer that particular question and providing them with further clarification on that item.

The procedures for the mailing of the second questionnaire were precisely the same as for the first, with one exception. Teachers who had not sent in the first questionnaire were mailed a modified second questionnaire that in fact was a combination of both questionnaires. The cover letter on this version of the questionnaire was tailored to that effect. The result of these efforts was that of the 67 teachers who were sent questionnaires, 61 teachers responded. Of those 61 teachers, 8 never completed the item on instructional practices, so that 53 teachers remained eligible to be included in the final analysis.

At the end of the school year, I submitted those 53 teachers' names to the school district's Office of Research, along with the specific class period of those teachers for which I needed rosters. The Office of Research requested from those schools class rosters for those classes. Two schools, which were asked to supply one class roster each, did not comply; the ultimate result was that class rosters for 51 teachers were returned. A clerk at the Office of Research extracted student data for those students listed in each of the 51 classes. These student data were aggregated into class data, which was the unit of analysis. The 51 teachers whose classes composed this data set came from 18 schools, the 20 schools whose teachers completed questionnaires minus the 2 schools with missing class rosters. Thus, 69% of the middle schools in this district

and 56% of the eighth-grade teachers (based on approximately 91 teachers) were ultimately represented in the student data analysis. The analytic sample consisted of 76% of teachers who were originally sent questionnaires.

Table 1 shows the means of relevant demographic characteristics for the 18 schools that were included in the analytic sample. These data were compared to the means of the same variables from the eight schools that were not included in the final sample. For the five independent-samples *t*-tests conducted, no significant differences were found among these demographic characteristics. This suggests that the 18 schools in the sample are likely to be representative of middle schools across the school district.

The MCLs were asked to name the specific mathematics courses offered to eighth-graders. The responses revealed four main categories: Math 8 (also called “General Math 8”), Pre-Algebra, Algebra I and Algebra II (also called Gifted and Talented). Occasionally, a school offered a course prior to Algebra I that by title did not exactly match one of those four categories. However, after discussing the nature of these courses with each MCL, such courses were categorized as either Math 8 or Pre-Algebra for this study. For example, one school did not offer Pre-Algebra, but instead offered Algebra IA, which was the first half of Algebra I extended over the course of the school year. For the purposes of this study, that course was considered to be Pre-Algebra.

Table 1
Sample Means (and Standard Deviations) for Middle School Demographic Characteristics for Schools in Analytic Sample or Excluded from Analytic Sample.

| Demographic | Mean (and SD) | |
|---|---------------------------------|----------------------------------|
| | Sample Schools <i>n</i> = 18 | Excluded Schools <i>n</i> = 8 |
| Enrollment | 924.3 (213.73) | 1019.5 (260.67) |
| Average daily attendance rate | 94.68 (1.05) | 94.81 (1.24) |
| Percent of students receiving free or reduced price meals | 27.66 (16.37) | 27.14 (15.51) |
| Mobility rate ^a | 20.85 (8.87) | 20.96 (11.92) |
| 6 th grade Mathematics CTBS ^b | 52.61 (19.75) | 52.75 (20.80) |

^aMobility rate is measured as the % entrant + % withdrawal. ^bMedian national percentile rank in the Comprehensive Test of Basic Skills.

The MCLs from each school were asked to furnish the teaching schedules of all eighth-grade mathematics teachers. The purpose of this was twofold: to determine the number of sections of each course offered at the schools, and to identify a particular section that each teacher would be asked to consider when filling out both parts of the questionnaire. Although the frequency of sections addressing the four courses was different, the goal of this study was to characterize instructional practice across all typical eighth-grade mathematics courses in the district. Thus, an attempt was made to obtain a reasonable sample size from all four courses, while also trying to ensure that each course was represented from each school. This was only possible when there were at least as many teachers as there were courses. Table 2 shows the result of this effort regarding representation of tracks in the analytic sample.

The table shows (a) that Pre-Algebra and Algebra I were offered at more than twice the rate of the other two courses over the district as a whole, and (b) that courses addressing lower-level mathematics were less represented in the final sample. This is particularly true for Math 8. One of the reasons for this is that fewer sections of those classes were taught and subsequently surveyed, so each section of Math 8 was associated with a larger potential percentage of response. The lower percentage reflects the sensitivity to sample dropouts, even when the reason for being excluded was *not* due to teacher non-response. In Math 8, one teacher who returned her questionnaire was not included in the data set because the district was unable to obtain her class roster. Another Math 8 teacher left the district halfway through the year. Thus, approximately 14% of the potential Math 8 response rate was not available for reasons not associated with the focus of this study. However, most of the missing classes were excluded due to non-response to the survey.

Table 2
Number of Sections Offered for Different Courses.

| Base | Course | | | |
|--|--------|-------------|-----------|------------|
| | Math 8 | Pre-Algebra | Algebra I | Algebra II |
| 18 Schools ^a | 28 | 68 | 73 | 31 |
| Sections surveyed ^b | 14 | 19 | 18 | 16 |
| Sections in analytic sample ^c | 8 | 13 | 16 | 14 |
| % Sections in final sample ^d | 57.1% | 68.4% | 88.9% | 87.5% |

^aTwo of the 20 cooperating schools did not provide section totals. ^b20 schools, 67 teachers surveyed. ^c18 of the 20 cooperating schools with student rosters for 51 teachers. ^dBased on the number of that section surveyed.

Mathematics in Arbor County School District Middle Schools

At a meeting with Ms. Byrd from the Office of Mathematics, I learned that there was an established lesson structure that the district and individual schools expected of their classrooms. That structure consisted of the following activities in sequence: drill/warm-up; teacher-directed instruction; guided practice; closure; assessment. This instructional model is traditional in nature and, according to the Office of Mathematics, not aligned with their philosophy. That is, the Office of Mathematics believed that the professional development they offered to their mathematics teachers stressed a lesson structure and teaching method that differed substantially from the one the district was suggesting.

Telephone interviews with each school's MCL provided specific information about the schools' mathematics programs. In particular these data included: types of eighth-grade mathematics courses offered, number of minutes in a mathematics period, textbooks used in each course, extra mathematics support offered to students, instructional resources available to teachers, number of eighth-grade mathematics teachers at the school, and student placement procedures.

Of the 18 schools in the study, 16 of them had schedules that accommodated mathematics class every day for between 45 and 50 minutes. One school had a 90-minute mathematics period every other day, and one school had a 65 minute-period four times per week.

In 10 schools, there were only three eighth-grade mathematics courses. Nine of those schools used the standard course titles of Pre-Algebra, Algebra I and Algebra II for their mathematics offerings. One school used a unique name to identify the low-

level course, but for the most part the content of that course seemed aligned with the Math 8 course described earlier. The other eight schools offered a fourth course that was usually called General Math 8, or Math 8. The schools offered varying numbers of sections of the different courses. Distribution of course offerings could be extreme. For example, one school had only one section of Pre-Algebra, eight sections of Algebra I, and five sections of Algebra II. Another school offered six sections of Math 8 and four sections of Pre-Algebra, but only registered two sections of Algebra I and one section of Algebra II.

The Algebra courses had been updated in recent years, and the district had every school using the same commercial textbook series for those courses, which they believed emphasized concepts over manipulation. Teachers were expected to move from the concrete to the abstract. Algebra I, in particular, had also been undergoing some modification in order to make it more congruous with the High School Assessment program. Changes that began during the year of this study, and were formalized the following year, resulted in more data analysis being included in the curriculum, along with more critical thinking and writing. Overall, the Office of Mathematics felt that these two courses were aligned with curriculum standards espoused by the National Council of Teachers of Mathematics. Those courses also had high stakes for students in that if they passed the final exam, they would get high school credit for the course. The final exams in both courses were district-wide, consisting of two parts. There were 30 multiple-choice questions worth a total of 70 points and an extended, constructed response worth 30 points. Calculators were not allowed on the multiple-choice questions. These questions were typical algebra problems, requiring

symbol manipulation and graphing sense. Part II of the district-wide final exam allowed the use of calculators for the four to six multi-part items. Those items were nearly all context-based, and for both courses, required students to recognize patterns, move between various representations of algebraic relationships, and write conclusions. To characterize the final examinations using vocabulary central to this study, Part I of the exams tested traditional material in a fairly traditional manner; Part II contained items that were consistent with instructional models more prevalent among reform instruction. Together, the two parts appeared to assess whether students could apply procedures in familiar contexts and in non-routine settings.

The Pre-Algebra and Math 8 curricula had not been reviewed or revised recently by the district's Office of Mathematics. In many schools, the distinction between those two curricula had been becoming more and more blurred. The Pre-Algebra course was originally designed to be just that – a course that would prepare students for algebra. However, the Office of Mathematics maintained that the course was often not as rigorous as it needed to be, in part because of the textbook many schools were using. Math 8 was designed to be a curriculum for students who were deemed below grade level. It was supposed to include some Pre-Algebra material with additional skill work and remediation. In some schools, teachers chose to enhance the curriculum of Math 8 to make it look more like Pre-Algebra. The book that was meant to be used for Math 8, called *Mathematics Plus* (1992), was actually also used at many schools that offered Pre-Algebra as their lower-level course. Ms. Byrd asserted that *Mathematics Plus* was a “creative curriculum guide” that was expected to help teachers approach old material from a new, usually contextualized and motivating standpoint. However, Ms. Byrd was

aware that many teachers were not teaching the course this way, but instead, focused on a skills-based curriculum. To reiterate, the distinction between Pre-Algebra and Math 8 was very ambiguous, and the course titles seemed to conjure different things depending on the school.

To further put these courses in perspective, it is useful to know in which courses students would be placed once they were in ninth grade. Math 8 students would likely enroll in a course called “Algebraic and Geometric Topics I” (AGTI), which is a course similar to Pre-Algebra. Eighth-grade Pre-Algebra students would go into either AGTI or Algebra I, depending on their teachers’ recommendations. Algebra I students would generally follow into Algebra II, and Algebra II students would be placed in one of two geometry courses, depending on teacher recommendations.

Students were placed into their eighth-grade courses based on seventh-grade test scores, grades, and predominantly the recommendation of their seventh-grade teacher. Several MCLs told me that the determination of whether a student was placed into Math 8 or Pre-algebra was essentially based on teacher recommendation. In at least one school, the Math 8 course was specifically for students who had not passed the state’s Functional Math Test (FMT). Eight schools said that whether or not students passed the state’s FMT determined their eligibility for Algebra I. That is, these eight schools said that students who had not passed the FMT were not eligible for taking Algebra I in eighth grade. The other ten schools said that non-passage of the FMT did not automatically disqualify a student from taking Algebra I. In some of these schools, there were so few students that had not passed the FMT by the end of seventh grade, that it

was a non-issue; in other schools, the FMT was not considered in course placement and there could be students in all courses who had not yet passed the FMT.

Four of the 18 schools also had a special program, called “Algebra with Assistance.” This meant that at the end of seventh grade, certain students were identified as “having potential” to pass Algebra I, if given extra assistance. The students identified as such were mixed into the Algebra I classes in eighth grade, but went for extra help in groups. The extra help took place usually five out of ten days during a special period. It was generally done with a different teacher than they had for the regular mathematics period and consisted of tutoring, or “coaching.” One school offered Algebra IA, which was the first half of the Algebra I course. Another special program occurred in one school that had no course below Pre-Algebra. All of their Pre-Algebra students had a separate “prep” class with a different teacher, in which they could prepare for the FMT or the state PA.

The MCLs were also asked to estimate the percentages of students in the lower-level courses who had already passed the FMT by eighth grade. Their responses further illustrate the severity of the differences in schools. In five of the schools, many or most Math 8 or Pre-Algebra students had already passed by the end of seventh grade, and nearly all had done so by the October administration of the test. In approximately seven schools, less than 20 % of students in the lower-level courses had passed the FMT, and in one school, almost none of those students had passed.

A follow-up interview question asked what kind of support the school offered for students who had not passed the FMT in eighth grade. Various programs were cited, and many schools had multiple supports. Of the 18 schools, 13 had pull-out programs,

usually scheduled around an up-coming administration of the test. Generally, any students who had not passed the FMT were rotated into a pull out section, with higher performing students entering the rotation sooner. Half of the schools offered after-school help; some of those programs were more formally structured than others. Two schools mentioned having a Saturday program. These programs were usually not conducted by the classroom teacher, but rather with the help of computerized instruction, resource teachers, special education teachers, aids, tutors, school volunteers, instructional assistants, or the MCL for that school. As mentioned previously, one school had a second mathematics course taken by all students in Pre-Algebra. This second mathematics course was for the sole purpose of preparing students for their two eighth-grade state assessments: the FMT and the PA.

When MCLs were asked about instructional resources that were available to eighth-grade teachers and students, replies were nearly uniform. All MCLs mentioned manipulatives, algebra tiles, scientific and graphing calculators, computers, and software. For the most part, it sounded as though both kinds of calculators were used most extensively, with the other resources used only moderately or not much at all, according to the individual teacher's discretion.

Instruments and Variables

Two instruments were used in this study. The first was a study-developed, two-part survey instrument (Appendices B and C) that supplied the teacher data; the second was the statewide assessment named in this study "PA."

Teacher Survey

Because a teacher's instructional model was a crucial variable in this study, the teacher questionnaire was developed after consulting several teacher surveys which included items addressing instructional practices (Herman, et al., 1994; Koretz, et al., 1996; Mayer, 1998; Raudenbush, et al. 1993; Shepard & Dougherty, 1991; Wiley & Yoon, 1995). The questionnaire also contained items that surveyed teachers' perceptions of their class' ability level, professional development exposure and participation, class environment, professional environment (administrator support and collegiality), test preparation pressure, knowledge of and attitude toward reform pedagogy, and the match between content taught and topics that could appear on the PA. The variable established to address the level of reform instruction employed by teachers used questionnaire items that addressed the frequency and duration of use of various instructional practices. This variable will be discussed in detail following a discussion of the questionnaire as a whole. Table 3 lists all of the variables assessed by the teacher survey, the location of these variables in the questionnaire, and the scoring process for each variable. Note that most variables were assessed with multiple sub-items answered on a seven-point rating scale. Typically scores were calculated by taking a simple average of the sub-items, except where noted in Table 3. Internal reliability was evaluated with Cronbach's alpha coefficient. Those coefficients, which ranged from a low of .66 to a high of .84 are listed in the table.

The scales for many of the items were labeled with descriptors only at the extreme ends and some of them included a center-position and two off-center descriptors. This was done to avoid too many descriptors that could be difficult to

distinguish or confusing to the reader. The odd-numbered scale was chosen to allow the respondent to choose the neutral position.

Table 3
Teacher Variables Assessed in the Questionnaire

| Variables | Location in Questionnaire | Scoring |
|---|---------------------------|---------------------------------------|
| Course name | Part 1, #1 | nominal |
| Professional Environment | Part 1, #5 | average score for a-d; $\alpha = .74$ |
| Class Problems | Part 1, #4 all | average score: a-j; $\alpha = .73$ |
| Class Diversity Level (teacher's perception) | Part 1, #2 | number selected |
| Class Ability Level (teacher's perception) | Part 1, #3 | number selected |
| Education | Part 1, #9,12 | qualitative, descriptive |
| Credentials | Part 1, #10 | Coded as 1, 2, 3, or 4 ^a |
| Professional development | Part 1, #13,14 | No. given, no. selected |
| Years teaching mathematics | Part 1, #11b | number given |
| Teacher control | Part 1, #6 | average of a-e, $\alpha = .66$ |
| Importance of reform pedagogy | Part 1, #7 left side | average of a-i; $\alpha = .73$ |

(table continues)

| Variables | Location in Questionnaire | Scoring |
|---|---------------------------|--|
| Preparedness for reform pedagogy | Part 1, #7 right side | average of a-i; $\alpha = .84$ |
| Familiarity with various documents/tests ^b | Part 2, #2 left side | number selected |
| Influence of various documents/tests ^b | Part 2, #2 right side | number selected |
| Conflict questions ^c | Part 2, #4 | composites from alpha factoring |
| Content taught ^d | Part 2, #1 | Separated into four groups |
| Level of Reform Instruction | Parts 1, # 8; Part 2, #3 | Ratio of reform practices to total reported class time |

^a1 = provisional or emergency; 2 = non-math. or elementary, or special education; 3 = elementary plus 18+ hours of mathematics; 4 = mathematics. ^bThese were treated as five separate variables – used in descriptive analysis. ^cThis question was slightly different for teachers of algebra vs. non-algebra. ^dUsed in descriptive analysis to compare coverage for different tracks of rational numbers, geometry and measurement, statistics and probability, and algebraic concepts.

Although this questionnaire was adapted from previous ones that had their own assessments of psychometric properties, as a modification of other instruments ordinarily reliability and validity should be re-established (Creswell, 1994). However, the notion of a reliability study that re-surveys a subgroup of the original sample with the entire survey was thought to be too burdensome on the teachers. Additionally, such a reliability check assumes that we want or expect consistency in how teachers answer the questions.

In the case of instructional practice, this assumption might be misplaced. That is, teachers might very well change their instructional strategies over the course of the school year for sound reasons. Thus, no such reliability check was done. Rather, in

order to ensure a more reliable picture of what teachers were doing in their classrooms, the survey was sent in two waves, once in November and subsequently in March.

The November survey contained items whose answers were more likely to be stable – teacher background and opinions about school and teaching situations that were more reflective of the teacher’s overall experience in their current position. This survey also asked the teachers about their instructional strategies. In March, teachers answered questions that were likely to yield more accurate answers toward the end of the year, such as feelings of pressure regarding the PA, content coverage, and again, instructional strategies. Thus, the questions addressing instructional strategies were answered twice, using exactly the same format of frequency of use and duration. There were a few exceptions to the instructional strategies questions being answered on two separate occasions by all teachers: Some teachers simply omitted that question on one or the other questionnaire; a few teachers only responded to the first survey; and some teachers who had not returned the November survey did respond to the complete two-part survey in March, but this altered questionnaire logically only included one set of the items on instructional strategies. The data from these teachers were retained in order to achieve the largest viable sample possible.

Teachers’ Instructional Style

The items on instructional strategies were combined to form a variable for the level of reform instruction. Since it is the central variable for this study, taking two “snapshots” of the teachers’ reported use of instructional strategies was thought to be more reliable. The assumption was that a particular teacher’s instructional style likely

does not vary much without intervention, although he or she might adapt strategies to particular content matter. Mayer (1999) showed that a composite measure of reform such as the one used in this study, rather than the individual instructional strategies, had good reliability based on surveying a sub-sample of teachers on two separate occasions and correlating the two composite measures of reform-strategy use. He found the correlation to be .69 ($p = .0013$) based on 19 teachers out of 20 that were sent a second questionnaire. The correlation from this study for 41 teachers who answered the instructional strategies question on both occasions was .789. At each instance, the scores were calculated as described below. Since there was no way to know if one score was more accurate than the other, or if the teacher truly changed instruction as the year progressed, the two scores were combined by taking an average to use as a final score on that variable. In the ten cases in which only one reform score was available, the missing score was imputed using the given reform score as a predictor in a regression analysis. Then the two scores, one supplied and one imputed, were averaged.

Observational data of classrooms of all responding teachers would be the ideal way to validate the instructional strategies portion of the survey. However, there was no practical way to accomplish this that would yield statistically sound measures of validity. There would need to be too many teachers observed over a lengthy period, which would not be permitted by the school district. A short discussion of the ability of survey data to capture classroom practice is discussed below with regard to a previous study. The present study included a qualitative component that described the instructional practices and classroom atmospheres of three teachers who were identified as exemplary teachers by administrators. Although a qualitative description of the

classrooms of only three teachers could not provide insight as to the validity of the survey in general, comparisons are made between the questionnaires of those teachers and the instructional practices seen during observations. A detailed description of this qualitative study is described in chapter 5.

Scoring and Validation

The instructional strategy component of the teacher questionnaire was used to give teachers a score on the level of reform approach. The survey asked about the frequency and duration of use of 24 instructional strategies. Of these strategies, 14 were coded as reform, and 10 were coded as traditional. The items themselves were adapted from many different surveys that attempted to assess the instructional model used in a classroom. The distinction of each item as being either a traditional or reform approach is consistent with the pedagogy espoused in NCTM's *Principles and Standards* (2000). The score for instructional model is actually a score on reform strategies, since it is a percentage of total reported class time that those reform strategies were reportedly employed. Those 14 reform approaches yielded an internal reliability coefficient of .71 for the surveys collected in this study.

The approach of addressing both frequency and duration of strategies is based upon Mayer's (1998) study. However, Mayer used a 5-point rating scale for duration, and then used a conversion to make a quasi-continuous scale. The survey used in this study asked the respondent to estimate the number of minutes per period directly, and so no conversion was necessary.

Table 4
Frequency of Instructional Strategy Use as Converted to Days in the School Year

| Descriptor | Conversion to days | Justification for conversion |
|-----------------------|--------------------|---------------------------------|
| < 3 times per year | 1.5 | Average of 1 and 2 |
| 3-9 times per year | 6 | Average of 3 to 9 |
| 1-3 times per month | 20 | 2 times per month for 10 months |
| 1 or 2 times per week | 55 | 1.5 days per week for 36 weeks |
| Nearly every day | 144 | 4 days per week for 36 weeks |

The score was computed as follows. First, the five-point scale for frequency of use was converted to the number of days of the school year by making the following correspondences, which were based on a 36-week, 10-month, 180-day school year. Table 4 lists the equivalences used to convert the response to a number of school days to each possible response for frequency.

The number of days was multiplied by the number of minutes of reported use. Inevitably, some teachers failed to provide either a frequency or duration for one or more of the items. Out of 48 total responses for duration and frequency, however, no teacher was missing more than six. These missing data were filled in with imputed numbers. Imputation on frequency, for instance, was carried out using all other filled-in frequencies as predictors in a regression analysis. The final score was an equally weighted composite variable formed by dividing the total time for all 14 reform strategies by the total number of minutes reported. This method was necessary because if the denominator were simply the total number of minutes in the school year, there could be an un-interpretable ratio of more than one. This could happen if teachers reported using several strategies at once, allowing for the possibility that the instructional minutes reported were more than the minutes in the school year. It is

noted however, that the measure as reported could be deceiving because, for instance, if there were traditional practices that the teacher used that were not listed, the teacher would have an inflated reform score. However, numerous surveys of instructional practices were examined; this survey included all of the strategies (reform and traditional) that could be located, as well as other items that might have been presumed to be subsets of items on previous questionnaires. Thus, the survey used for this study presents a nearly exhaustive listing of possible instructional practices that teachers may use in a mathematics classroom. The score, which is a percentage of time that the reform strategies on the survey were used, was a key variable for several analyses.

Mayer (1998) attempted to validate the items by correlating the survey reform score with a reform score based on three observations for a subset of teachers. He found that the survey accurately captured the variation in reform usage, although a qualitative description revealed that some reform approaches were used only superficially. That is, some teachers who scored high on the reform scale did so because they used some “reform” methods only on the most superficial level. The items used to assess reform instruction in my study were modifications of items from previous surveys in an attempt to mitigate these pitfalls to some extent. Thus, an item such as “calculator usage,” which could be used in a very traditional manner or as a reform strategy, was replaced with the more specific items: “use calculators or computers to practice skills” and “use calculators or computers as a tool (e.g., spreadsheets) or to explore patterns.” The former item was classified as traditional, whereas the latter was classified as reform. Although the items were carefully constructed to avoid some obviously ambiguous (with regard to reform or traditional teaching) items, survey data can not by any means

provide perfectly accurate accounts of classroom practices that can really only be measured precisely with intensive observational periods.

Because it is inadequate to conduct a validation study by observing teachers during only two or three visits, I chose instead to observe only three teachers, but to observe them repeatedly over a longer period. All of the teachers were observed over a time span of 3-5 school weeks; one teacher was observed on 8 days, and the two remaining teachers were observed on 13 days each. The purpose of this component of the study was simply to describe the instructional practices and classroom atmosphere of exemplary teachers of high and low ability students, as defined by their course. Although one could not generalize from only three teachers, the self-reported scores of these teachers were for the most part consistent (in an ordinal fashion) with the instructional styles observed. That is, the teacher whose style was observed to be most in line with reform practices had the highest reform score of the three teachers, and so forth. In chapter 5, I provide a more in-depth discussion about the match between these teachers' self-report data from the survey and their instruction as observed in their classrooms.

State Performance Assessment

The state performance assessment (PA) is a criterion-referenced “authentic” assessment that was administered to nearly all¹ eighth-grade students in the state during the first week of May. State documents on the assessment contend that the test measures “higher-order thinking processes and the application of knowledge and skills to real

¹ Some students with disabilities and some ESL students may be exempted from the state performance assessment.

world or authentic situations” (MSPAP, p. 2). The test is used to measure school performance, not individual student achievement. Students are tested for nine hours over five days, with mathematics being one of six content areas tested.

The assessment consists of tasks that might comprise several items or activities that are related to one theme. Tasks may encompass more than one content area. Activities may be group or individual activities; they may require observation, reading or hands-on engagement. Questions require students to either solve a problem, make a decision, or give an explanation. Some calculator use is allowed, as well as the use of manipulatives (Bond, Roeber, & Connealy, 1997).

The reported alpha coefficient for the mathematics total score on the PA is .85. For the area of mathematics process the coefficient alpha is only .65 because there are fewer items on this test. However, for the purpose of school-wide instructional decision-making, the reliability data is considered acceptable to the state. This study included only the mathematics total score. Tasks on the assessment undergo multiple reviews prior to implementation. During the review process, tasks are checked for content and face validity, and that they are assessing knowledge described in the state’s learning outcomes document (MSPAP, 1998). The testing contractor produces scale scores that range between 350 and 700, with a mean of approximately 500 and a standard deviation of approximately 50.

Other Important Measures

Other data on school demographics were also collected from the school district: school mobility rate, percent of students on free or reduced lunch, school size (number of students), school attendance rate, and the schools’ median national percentile rank on

the mathematics portion of the Comprehensive Tests of Basic Skills (CTBS). These data were used in a descriptive analysis of the schools, and as controls in the analysis on the link between reform teaching and PA achievement.

Class data, averaged across students, were collected for each class about which a teacher was surveyed. Those data were: class size, percent of students in the class on free or reduced lunch, sixth-grade CTBS scores for total language, reading, and mathematics, and PA scale score. An additional variable was the percent of students in each class who had passed the FMT by the middle of their ninth grade, which was the only FMT information available at the time that the data were accessed. Because the district did not have on record grades for students prior to high school, the sixth-grade CTBS scores for students were obtained as a substitute for seventh-grade grade-point average. The CTBS is a norm-referenced assessment program that provides information about students' grade-level performance, as compared to national norming samples. Sixth-grade CTBS mathematics, language and reading scores were obtained for use as a measure of student achievement. Analyses showed that all three scores were highly correlated, so only the mathematics CTBS scores were used as a control variable.

In some cases, student CTBS data were missing. T-tests revealed that in five classes, PA scores were significantly different for students whose CTBS scores were or were not available. Thus, rather than excluding these students from class averages, student CTBS scores were imputed before obtaining class averages. These imputations were carried out using PA scores and other-subject CTBS scores when available.

Data Analysis

One of the concerns of this study is opportunity for *all* students to learn, hence comparisons were made across the course levels for many variables of interest. However, upon examining the two courses, Math 8 and Pre-Algebra, it appeared that the distinction between the two courses was very unclear. The specialist in the district Office of Mathematics agreed with my suggestion that for the purposes of this study these two courses should be collapsed into one category. In order to validate this decision, I conducted some preliminary analyses to examine whether there were any notable differences between these two lower courses in terms of student background or between certain class-level variables as reported on the teacher questionnaires.

Concerning variables related to student background, there were no significant differences between the percent of students on free or reduced lunch or on the CTBS mathematics scores for students in Math 8 classes versus students in Pre-Algebra classes. Furthermore, there were no significant differences found for these two courses in terms of the teachers' perceptions of the ability of the classes, professional development, years teaching mathematics, and the reported amount of preparation students in those classes would receive in the four learning outcomes areas. Thus, for the remainder of analyses, data from Math 8 and Pre-Algebra were combined into one lower-level course. In the descriptions and analyses that follow pertaining to distinctions between *three* courses, "Pre-Algebra" is used to categorize the course below Algebra I and Algebra II.

Once the two lower courses were collapsed into one category, preliminary analyses focused on describing important characteristics of the three course levels. The

teacher survey assessed not only background teacher characteristics, but also teachers' perceptions of their students and environment, content coverage, and teachers' impressions regarding test pressure and alignment of state tests and their courses. For purposes of description, the reported results begin with an overview of those variables ascertained from the teacher survey and broken down by course. Where appropriate, frequency tables are used to illustrate differences among the courses, as in the case of the credentials held by the teachers.

Prior to conducting any regression analyses, I found Pearson correlations between all of the aforementioned variables. First, the correlations offered further descriptive interest. Second, the correlations provided a glimpse at whether there were highly correlated variables that could possibly cause a washing-out effect when regressing reform usage on them or spurious relationships that cause exaggerated correlations between a predictor and the dependent variable. Thus, the correlations helped to inform the regression analyses.

One of the main research questions concerned characterizing the instructional models used in the three different courses. Therefore, the most commonly used teaching strategies, as reported by the teachers, were summarized and graphically displayed for ease of comparisons across the courses.

After thoroughly examining the various characteristics of the teachers and their instructional strategies, I performed a regression analysis to determine useful predictors of amount of reform instruction, as operationalized in the way described previously. There are numerous teacher variables that could be thought to predict reform usage, including attitude toward and preparation to use reform pedagogy, familiarity with

NCTM or state documents, professional development activity, classroom environment, conflict among curriculum and PA preparation, and the ability of the class, as perceived by the teacher. However, a limiting sample size compelled me to take a more conservative approach as far as choosing predictors in a regression analysis for reform instruction. In being selective of the predictors, I first noted that almost none of the variables that could be hypothesized to be determinants in the amount of reform usage were significantly correlated in either direction with reform. This indicated that it was unlikely that relevant predictors were being washed out by opposite correlations with other possible predictors.

Furthermore, research has pointed to a disconnect between attitudes toward and knowledge of reform, and teaching practices. For example, Guskey (1994) pointed out that even teachers who had a positive attitude toward performance assessment were unable to readily adapt their teaching strategies to align with the assessments. In a study on the Third International Mathematics and Science Study (TIMSS) by Stigler, Gonzales, Kawanaka, Knoll, and Serrano (as cited in Jacobs and Morita, 2002), the authors found that even when teachers were familiar with various reform documents, they still maintained traditional teaching practices. In light of these studies and other analyses already completed in this study, only a small group of variables was entered into the regression analysis for reform usage.

The third research question, which concerns teachers' reform score and class achievement on the PA, involved student variables. Because the reform score is a classroom level variable, all student variables were considered at the class level by taking an average, hence the collection of only aggregated data for each class. Since

classes are nested within schools and schools might act autonomously with respect to key aspects of their eighth-grade program (length of period, extra remediation, grouping practices), ideally, analysis would have been conducted at the school level. This would have eliminated the problem of inter-dependence among teachers and classes from the same school. However, there were simply not enough schools in the district to suffice. Even if another district had been added to include more schools, this would have introduced the additional layer of district-level dependence among schools within the same district, to contend with. There were also not enough teachers within any one school to merit such a hierarchical analysis, nor were there enough teachers to do an intraclass correlation analysis that might have shown negligible differences within schools. Another way to eliminate the school effect would be to partial out school identities by mean centering the data. However, this technique would have been inappropriate for most variables since I was not interested in a class or teacher's score only relative to their own school. Ignoring school differences was certainly a limitation of this study, but many important characteristics of the mathematics programs were essentially identical across the district, such as the length of a class period and placement procedures. Thus, the data were analyzed with the classroom as the unit of analysis. In analyses that used student data, all data were weighted by relative class size.

When this study was initially conceived, I hypothesized that a mediation model might best describe the relationships among course level, reform instruction, and achievement on the PA. Figure 1 illustrates this hypothesized relationship. In this model, the course level predicts the amount of reform instruction, which in turn, predicts the level of achievement on the PA. In other words, reform instruction acts as a

mediator variable to the extent that it accounts for the relationship between the course level and the achievement on the PA (Baron & Kenny, 1986). Other conceptually related variables, such as the average sixth grade CTBS mathematics score and the percent of students on free or reduced lunch, would be included in the analysis as controls, but are not included in the pictorial for ease of description.

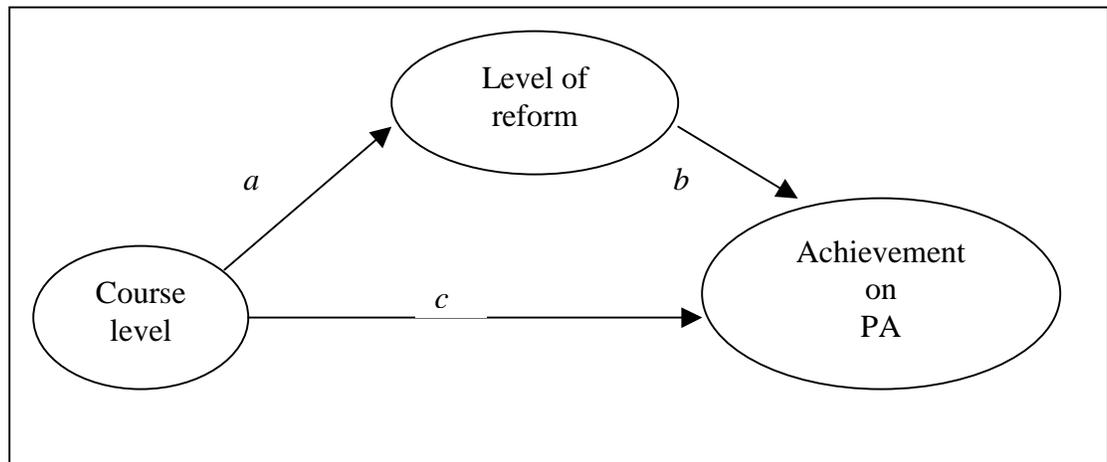


Figure 1. Mediational model, in which level of reform is the mediator.

In order for the level of reform to function as a mediator in the relationship depicted in Figure 1, several relationships must be established: Path *a* must be significant, indicating that variation in the course level accounts for variation in the amount of reported reform usage; Path *b* must be significant, indicating that variation in reform usage accounts for variation in achievement on the PA above and beyond course level; When both course level and reform are included as predictors for achievement, a previously significant relationship between course level and achievement level is substantially diminished. That is, Path *c* should be close to zero when both predictor variables are included.

I also tested a second model of the relationship between the course level, reform usage, and PA achievement. The second model considers the possibility that an interaction, or moderator, effect is at work influencing the relationship between the relevant variables. That is, it is conceivable that reform might predict achievement on the PA, but only as moderated by course level. Figure 2 shows a pictorial view of this model, again leaving out other possible control variables for ease of description. If the

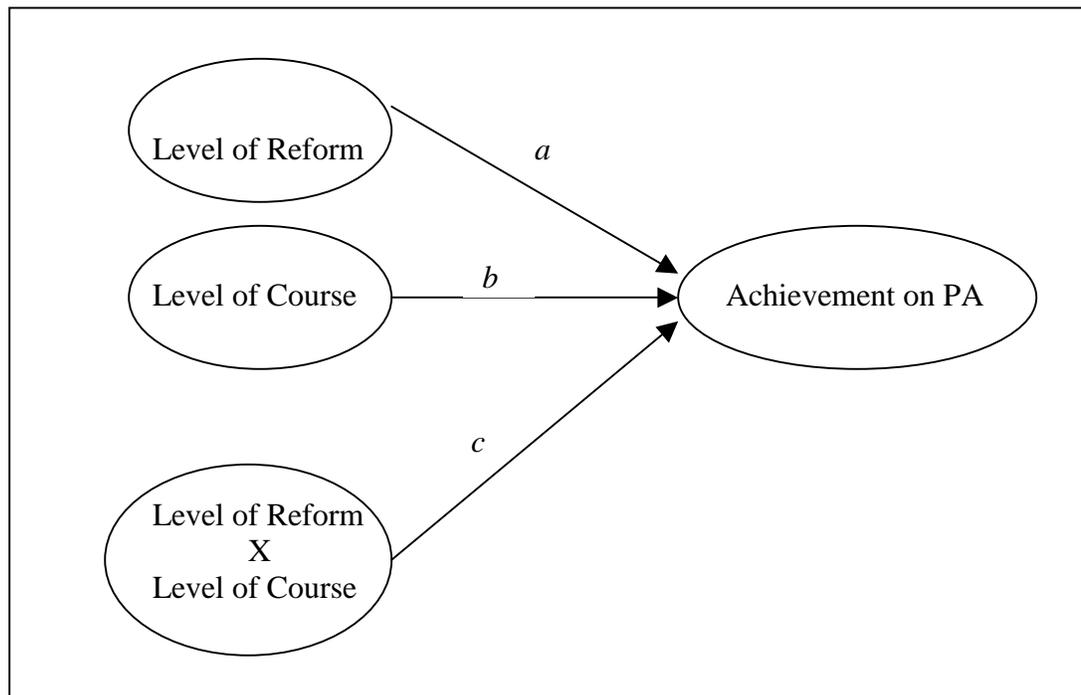


Figure 2. Moderator model, in which course level is the moderator.

moderator hypothesis holds, Path c , the path for the interaction term, would be significant. This would indicate that the relationship between reform and achievement on PA is different for different course levels. For example, it could be the case that for Pre-Algebra, the amount of reform usage is not predictive of PA achievement, whereas for Algebra II, level of reform *does* have predictive power for PA achievement.

For all of the analyses discussed above, level of course was operationalized using two dummy variables. A class either had Algebra II status, or it did not; a class either had Algebra I status, or it did not. If a class had neither Algebra II status, nor Algebra I status, it was necessarily a Pre-Algebra class. The interaction term in the moderator scheme was operationalized as the product term of the two relevant variables.

A summary of the main and subsidiary research questions, along with the analyses that were conducted is provided in Table 5.

Table 5
Research Questions, Sources, and Analysis

| | Research Question | Data Source | Analysis |
|----|--|---|---|
| 1. | Is there a difference among the courses in terms of the extent to which teachers have covered material from topics in the state's learning outcomes? | Survey, Part 2: Content coverage | ANOVA |
| 2. | Do teachers in different courses feel different amounts of conflict over teaching their curriculum and teaching toward the PA? | Composites from alpha factoring items from Survey, Part 2, #4. | ANOVA |
| 3. | What factors related to teacher, classroom or student characteristics influence the amount of reform instruction that teachers perceive they implement? | Survey | Multiple Regression Analysis with reform instruction as the dependent variable |
| 4. | Does the amount of reported reform instruction differ for 8 th graders in different academic courses? In what ways do instructional practices differ for 8 th graders in different academic courses? | Survey – instructional strategies | ANOVA |
| 5. | Is there a relationship between reported level of reform instruction and students' performance on the PA? | Survey; District supplied student data | Series of Multiple Regression Analyses |

CHAPTER 4: RESULTS

This chapter begins with descriptive analyses in order to (a) further delineate differences among the three levels of courses: Pre-Algebra, Algebra I, and Algebra II and (b) give an overall picture of relationships among the variables ascertained from the teacher survey. Since it had already been determined that Math 8 and Pre-Algebra were not easily distinguishable across schools, those two courses were collapsed into one, which is referred to as Pre-Algebra in the tables in this chapter. Following those descriptive analyses, I present results regarding the instructional models that teachers perceive they employ. Finally, results from the regression analyses concerning reform instruction and achievement on the PA are provided.

Descriptive Analyses

Teacher Background Variables

One notable difference among the three courses was the level of certification that the teachers possessed. Table 6 lists the frequencies for the four levels of certification into which the data were categorized. Those levels are hierarchical in terms of what kind of education stakeholders would ultimately want teachers of mathematics to have. The most obvious discrepancy is that whereas Algebra II teachers were all certified in mathematics, many of the teachers in the lowest course had neither secondary/middle mathematics certification, nor had they completed at least 18 credits of mathematics coursework.

Table 6

Frequencies of Certification Types of Teachers in Three Levels of Courses

| Certification Level | Course level | | |
|---|---|----------------------------|-----------------------------|
| | Pre-Algebra <i>n</i> = 21 ^a | Algebra I <i>n</i> = 16 | Algebra II <i>n</i> = 14 |
| Provisional or Emergency | 2 | 1 | 0 |
| Non-mathematics ^b or elementary or special education | 6 | 1 | 0 |
| Elementary +18 ^c | 1 | 4 | 0 |
| Mathematics (secondary/middle) | 11 | 10 | 14 |

^aOne data value is missing from this category that combined Math 8 with Pre-Algebra, so that frequencies sum to 20. ^bNon-mathematics refers to certification in secondary/middle in a field other than mathematics. ^cCertified in elementary education with 18 or more semester hours of mathematics.

Table 7 provides descriptive statistics for other teacher background characteristics. Teachers responded to a question about recent credits by recording the actual number of college credits (beyond requirements for initial certification) they earned from formal courses in mathematics or mathematics education. Professional development referred to professional meetings, workshops and conferences, but not formal credit-bearing college courses. In order to respond to the professional development question, teachers selected from among a 6-point scale, corresponding to an amount of in-service education. An ANOVA indicated no significant differences between teachers in the three courses in terms of the three background characteristics: recent credits, professional development, and years of experience teaching mathematics.

Table 7

Teacher Background Characteristics for three course levels

| Variable | Mean | SD | Min. | Max. |
|--------------------------------------|------|------|------|------|
| Recent Credits | | | | |
| Pre-Algebra | 3.40 | 8.11 | 0 | 36 |
| Algebra I | 2.53 | 6.33 | 0 | 24 |
| Algebra II | 2.79 | 2.75 | 0 | 9 |
| Professional Development | | | | |
| Pre-Algebra | 3.81 | 1.21 | 2 | 6 |
| Algebra I | 4.19 | 1.52 | 1 | 6 |
| Algebra II | 4.07 | 1.27 | 2 | 6 |
| Years Teaching in Grades 7-12 | | | | |
| Pre- Algebra | 10.1 | 10.1 | 1 | 31 |
| Algebra I | 10.6 | 9.3 | 1 | 31 |
| Algebra II | 14.4 | 8.4 | 3 | 30 |

Teachers' Perceptions of Classroom Characteristics

Unlike the teacher background characteristics, there were significant differences among the three courses for variables that measured characteristics of the students, as perceived by the teachers. Teachers rated the ability of their class on average from 1 (*generally low in ability*) to 7 (*generally high in ability*). Teachers rated the diversity of their class from 1 (*students are very similar in ability*) to 7 (*students are very diverse in ability*). The score for a variable titled “class problems” was ascertained by forming a

composite score for ten items, among them: student interest in mathematics, effort needed to maintain discipline, sufficient parental support, and interruptions for announcements or school activities. The 7-point scale for each item was labeled with two descriptors, 1 (*serious problem*) and 7 (*not a problem*). This scale was established as the reverse order of what might be expected only because other questionnaire items near it had the positive reactions at the upper end; thus, the order of the scale for class problems was established to be consistent with those. Table 8 lists these aforementioned variables, along with their means and standard deviations. ANOVA was used to determine omnibus F tests of differences among means. With only three groups, Fisher's least significance difference (LSD) test provided results for post hoc comparisons, based on an optimal amount of power while maintaining familywise error rates (Hancock & Klockars, 1996).

Among the three levels of courses, all of the means for teachers' perceptions of their class' abilities were significantly different from one another. This coincides with the reasoning behind offering different levels of eighth-grade mathematics courses in the first place. The differing perceptions of class diversity were somewhat more interesting. The Algebra II classes had the least amount of diversity, the mean of which was significantly different from the mean for diversity in Algebra I, where teachers reported perceiving their classes as being the most diverse. Algebra II in eighth grade is part of the gifted and talented program. Thus, students in those classes have in some way "proven" themselves academically and are likely to be more similar in ability. The fact that the Algebra I course reported the most diversity could be explained by the current trend to teach Algebra I to more students earlier than ever before. Indeed, as

Table 8
Comparisons of Means for Classroom Characteristics

| Variable | I | J | I-J | Standardized Effect Size |
|-----------------|-------------------------------------|---------------------------------------|--------|--------------------------|
| Class Ability | Algebra II $\bar{x} = 5.89(.96)$ | Algebra I $\bar{x} = 4.19(.98)$ | 1.7* | 1.30 |
| | Algebra I | Pre-Algebra $\bar{x} = 3.07(1.66)$ | 1.12* | .86 |
| | Algebra II | Pre-Algebra | 2.82* | 2.16 |
| Class Diversity | Algebra I $\bar{x} = 4.56(1.63)$ | Pre-Algebra $\bar{x} = 3.96(1.24)$ | .6 | .42 |
| | Algebra I | Algebra II $\bar{x} = 3.00(1.36)$ | 1.56** | 1.11 |
| | Pre-Algebra | Algebra II | .96 | .68 |
| Class Problems | Algebra II $\bar{x} = 5.26(.78)$ | Algebra I $\bar{x} = 4.22(.90)$ | 1.04* | 1.17 |
| | Algebra II | Pre-Algebra $\bar{x} = 4.02(.94)$ | 1.24* | 1.40 |
| | Algebra I | Pre-Algebra | .2 | .23 |

Note. Numbers in parentheses are standard deviations. Fisher LSD comparisons were made.
 * $p < .05$. ** $p < .01$.

described in the previous chapter, several schools even had special programs to encourage weaker students to enroll in Algebra I where they would be provided the support of extra assistance. These students ordinarily would not have qualified for Algebra I. The fact that Algebra II classes reported significantly less classroom challenges than in the other two courses is consistent with the belief that those students might be more motivated and academically focused than middle-school students in non-GT courses are often perceived.

Course Content

On the questionnaire sent in March of the study-year, teachers responded to a series of questions that pertained to content coverage in the surveyed class. The nine items listed were broad mathematics topics outlined in the State Learning Outcomes for the eighth grade. Teachers circled a response from 1 (*no coverage at all*) to 7 (*thorough coverage*). Those nine items were grouped into four categories: rational numbers; geometry and measurement; statistics and probability; and patterns, functions and algebra.

Using ANOVA, followed by Fisher's LSD, significant differences in content focus were found between some of the courses for all of the categories except statistics and probability. Across all courses and all content areas, mean score coverages ranged from a low of 4, which was the mean coverage of geometry and measurement for Algebra II, to a high of 6.3, which was the mean coverage of patterns, functions and algebra in Algebra I. Table 9 lists the significant mean differences in content coverage between the courses. The fact that coverage of rational numbers was reported as more thorough for Pre-Algebra than for either of the other two courses is not surprising, since that course had the most students who had not yet passed the FMT, which includes items addressing operations with rational numbers. On a related note, another notable significant difference that appeared among the three courses was the teachers' responses to the survey question regarding the extent of influence that the FMT had on their teaching. The 5-point scale for that question ranged from 1 (*not at all*) to 5 (*very much*). The mean for Pre-Algebra teachers was 1.49 and 2.16 points higher than that of Algebra

I and Algebra II, respectively. Those differences were significant at $p < .05$ and $p < .01$, respectively.

Table 9
Differences in Content Coverage for Three Courses

| Content Area | I | J | Mean Difference I - J | Effect Size |
|---------------------------------|-------------|-------------|-----------------------|-------------|
| Rational Numbers | Pre-Algebra | Algebra II | 1.75** | 1.35 |
| | Algebra I | Algebra II | 1.29* | .99 |
| Geometry and Measurement | Pre-Algebra | Algebra I | .88* | .71 |
| | Pre-Algebra | Algebra II | 1.43** | 1.15 |
| Patterns, Functions and Algebra | Algebra I | Pre-Algebra | .88* | .82 |

Note. Fisher LSD comparisons were made.
* $p < .05$. ** $p < .01$.

Since the FMT is a basic skills test emphasizing fractions, decimals, and percents, it follows that teachers who are influenced by it would spend more time on rational numbers. Geometry and measurement were part of the Pre-Algebra curriculum, but not prevalent at all in algebra courses, hence the significant differences in that category. Geometry and measurement items are also represented on the FMT.

Curriculum and Testing Conflict

One of the contentions at the start of this study was that teachers might feel conflicted in trying to cover their course content and to simultaneously prepare their students for the state PA. In particular, it was thought that perhaps the lower-level classes were overly influenced by the need to prepare for the FMT, to the exclusion of preparation necessary for the PA, a very different kind of test. Furthermore, Algebra I and II teachers knew their students faced a district-wide final exam that, if passed, granted high school mathematics credit. If teachers felt that they needed to cover the algebra content in a more traditional manner because of the final exam, those instructional practices and emphases might not be consistent with the state PA. It was further hypothesized that some teachers' preferred method of instruction might not be consistent with expectations for PA preparation. A series of questions on Part 2 of the survey dealt with these issues.

The survey questions submitted to the teachers of Math 8 and Pre-Algebra were slightly different than those submitted to the algebra teachers, as the questions were closely tailored to the respective teachers' situations. However, most of the questions for one group of teachers matched analogous questions for the other group. For example, the algebra teachers responded to the statement, "I feel that the pressures of preparing students for the [state PA] and the algebra final give me mixed messages about how to teach my class." The teachers of Math 8 and Pre-Algebra responded to the analogous statement, "I feel that the pressures of preparing students for the [state PA] and the [FMT] give me mixed messages about how to teach my class." Those two questions will be referred to as the variable, "mixed messages."

Several of the items were examined individually for comparison across course levels for descriptive purposes. One of those items was the “mixed messages” item, explained above. Another, referred to as “PA/teaching fit,” captured responses to the question, “Preparing my students for the [state PA] allows me to teach the way I want to.” A third item, referred to as “higher-order teaching,” captured responses to, “In this class, I have time to cover the basics of the course, and still expect students to engage in higher-order thinking.” Table 10 presents the means and standard deviations for the three courses. When the analysis for this series of questions was undertaken, scores for several items were reversed for the purposes of directionally aligning all items. However, means reported have been reversed back where necessary for ease of interpretation. Thus, on a 7-point scale, higher numbers represent more mixed messages, better PA/teaching fit, and more engagement in higher-order thinking.

A one-way ANOVA was completed to determine if there were significant differences among the responses. Fisher LSD post hoc tests were employed to determine where differences occurred. There were significant differences found among the scores for PA/teaching fit between Pre-Algebra and both Algebra I and Algebra II courses at the $p < .05$ level. The difference was more pronounced between Pre-Algebra and Algebra I, with a mean difference of 1.73 on a scale of 1 to 7. The larger mean for Pre-Algebra indicates that those teachers felt that preparing their students for the State PA was more aligned with their preferred style of teaching. According to the district specialist who described the courses to me, this might be considered surprising, since it was her belief that teachers of the lower-level course were not generally teaching those

courses in as non-traditional a manner as they might have wanted, nor as one of the predominant book series presented it. However, it might also reflect the fact that

Table 10
Differences in Perceptions of Conflict by Course

| Conflict Variable | I | J | I – J | Effect Size |
|-----------------------|---------------------------------------|--------------------------------------|-------|-------------|
| Mixed Messages | Pre-Algebra $\bar{x} = 3.95(1.66)$ | Algebra I $\bar{x} = 5.00(1.51)$ | 1.05 | .61 |
| | Algebra II $\bar{x} = 3.92(2.02)$ | Algebra I | 1.08 | .63 |
| | Algebra II | Pre-Algebra | .36 | .21 |
| PA/Teaching Fit | Pre-Algebra $\bar{x} = 4.33(1.49)$ | Algebra I $\bar{x} = 2.60(1.40)$ | 1.73* | 1.10 |
| | Pre-Algebra | Algebra II $\bar{x} = 2.92(1.88)$ | 1.42* | .90 |
| | Algebra II | Algebra I | .32 | .20 |
| Higher-Order Thinking | Pre-Algebra $\bar{x} = 4.33(1.65)$ | Algebra I $\bar{x} = 4.00(1.89)$ | .33 | .18 |
| | Algebra II $\bar{x} = 5.50(1.93)$ | Pre-Algebra | 1.17 | .65 |
| | Algebra II | Algebra I | 1.50 | .83 |

Note. Numbers in parentheses are standard deviations. Fisher LSD comparisons were made.

*p < .05. **p < .01

the algebra teachers did not feel that the content they were required to cover could have been taught in a mode consistent with the PA.

Furthermore, alpha factoring was undertaken to determine a reasonable composite variable that would capture an underlying factor representing conflict (or lack thereof) over teaching to prepare students for the PA while teaching to cover

course content. From the Pre-Algebra group, a factor made up of six of the ten items emerged, yielding an alpha of .8188 ($n = 21$). A similarly interpretable factor emerged from both algebra groups, with an alpha coefficient of .7795 ($n = 27$) when both algebra levels were factored together. However, the factor from the algebra group used only five items. In order to have a factor that would be most uniform in meaning, I used only four of the items from each factored list that had nearly identical statements in both groups. For each teacher, the scores on those four items were averaged to form a composite factor. Table 11 lists the items that made up that factor.

Considering the direction of the response scale, the factor might be described as “accord” regarding teaching for the state PA and covering their respective curricula.

Table 11
Alpha-Factored Items from Questionnaire, Part 2 that Made up Accord Factor
 Pre-Algebra classes^a

1. I do not have any difficulty in simultaneously preparing students for both the [FMT] and the [PA].
2. Spending class time on [PA]-type activities also helps students learn some “basics.”
3. I spend less time working [PA]-type activities because many students still need to learn the basics.*
4. I feel that the pressures of preparing students for the [PA] and the [FMT] give me mixed messages about how to teach my class.*

(table continues)

Algebra classes^b

1. I do not have any difficulty in simultaneously preparing students for both the algebra final exam and the [PA].
2. Spending class time on [PA]-type activities also helps students learn some algebra concepts or skills.
3. If I spend time preparing students for the [PA], I'm afraid they would not learn enough algebra content.*
4. I feel that the pressures of preparing student for the [PA] and the algebra final give me mixed messages about how to teach my class.*

^aAlpha coefficient for Pre-Algebra for these four items was .7229.

^bAlpha coefficient for algebra classes for these four items was .6872.

*Scores for these items were reversed for analysis and interpretation of final factor.

That is, lower scores represent less accord, or more conflict, whereas, higher numbers represent better accord, or less conflict. Table 12 provides descriptive statistics for this composite variable. The omnibus F test did not reveal any mean differences. The largest standardized effect size for the differences in means occurred between the Pre-Algebra and Algebra I courses and was .63. The effect sizes for the other two comparisons were half that.

Table 12
Means and Standard Deviations for Accord Factor

| | Mean | SD |
|-------------|------|------|
| Pre-Algebra | 3.62 | 1.36 |
| Algebra I | 2.83 | .99 |
| Algebra II | 3.23 | 1.36 |

Relationships among Survey Variables

Another preliminary analysis focused on the correlations among the variables relevant to this study. Table 13 lists the variables that had significant correlations at the .05 level.

Some of the relations noted in the table deserve special attention for their status as either confirming relationships that might have been presumed, or for their surprising dissonance with the same. The ability variable measured teachers' responses to the question, "When you think about this particular class, what is your impression of their ability level on average?" The scale had seven points, on which 7 = "generally high in ability." It is not surprising that this correlated highly with the PA score, if the PA score is indeed a measure of some kind of mathematical proficiency, and the teachers' perceptions of their class' ability is on target. Also, an unfortunate, yet common correlation was affirmed with this data: the more students on free or reduced lunch in a class, the lower the PA score.

Slightly more interesting is the high correlation between familiarity with and influence of NCTM standards. It seems that for these teachers, to know NCTM is to embrace NCTM, at least in their perception. However, average reform score did not correlate with either of those variables. This is surprising because the reform score was meant to be a measure of the amount of reform instruction – instruction commensurate with NCTM ideology – that teachers use. Teachers responded to the question about how much NCTM influences their teaching on a 7-point scale, ranging from 1 (*not at all*) to 7 (*very much*). It was similarly curious that the importance of reform variable did not correlate with influence of NCTM standards or average reform score. Importance of

Table 13
Significant Correlations among Relevant Variables

| Variable 1 | Variable 2 | Correlation |
|---------------------------------|-----------------------------------|-------------|
| PA scale score | Class ability | .733 |
| Familiarity with NCTM standards | Influence of NCTM standards | .788 |
| PA score | % Class on free or reduced lunch | -.734 |
| Certification | Class ability | .556 |
| Class problems ^a | Ability | .488 |
| Class problems ^a | Diversity | -.433 |
| Class problems ^a | % Class on free or reduced lunch | -.371 |
| Familiarity with NCTM standards | % Class on free or reduced lunch | -.492 |
| Years teaching | Influence of NCTM standards | .418 |
| Influence of NCTM standards | % Class on free or reduced lunch | -.550 |
| Importance of reform | Accord composite ^b | .438 |
| Years teaching | Certification | .475 |
| Years teaching | % Class on free or reduced lunch | -.498 |
| Certification | % Class on free or reduced lunch | -.471 |
| PA scale score | Certification | .544 |
| Class size | Diversity | .383 |
| PA scale score | % School on free or reduced lunch | -.359 |
| Average reform score | Years teaching | .472 |
| Average reform score | % School on free or reduced lunch | -.356 |

^a “Class problems” was scored on a reverse-order scale, so higher numbers represent fewer problems. ^b “Accord” refers to the composite score that resulted from alpha factoring. This variable represents the factor described previously as ease of teaching for the state [PA] while covering the respective curricula.

reform was a variable composed by averaging the ratings teachers gave to each of nine fundamental reform notions, regarding their importance in education. Teachers may perceive that NCTM influences their instruction, when their actual practice does not demonstrate that perceived influence. Furthermore, teachers might believe that these teaching ideals are in fact important, but still not be capable of realizing them in their classrooms. However, it remains curious that teachers' perceptions of NCTM influence and their beliefs about the importance of reform teaching were not correlated. It is interesting to note however, that the variable, importance of reform, was correlated with the accord factor. That is, teachers who felt more at ease covering their course content and preparing their students for the PA exam also believed more in the importance of reform ideals.

At first glance the correlations reported that involve the problems variable might seem surprising. However, the scale for those items was reversed, so that higher scores meant fewer problems. Thus, the pattern in the table shows that these teachers reported that higher ability students, classes that were less diverse, and classes that had fewer students on free or reduced lunch had fewer problems. Also, classes with higher scores on the PA had fewer problems.

Another set of correlations sheds some light on the characterization of which teachers are in which classrooms. The higher the percentage of students on free or reduced lunch in a class, the lower the number of years of teaching experience and the lower the certification level of the teacher, with these teachers reporting less familiarity with, and less influence by, the NCTM standards.

The higher the certification level of the teacher, the more years of teaching experience, the higher the perception of the class' ability, and the higher the PA score was for that class. Finally, the more years of teaching experience reported, the more influence of NCTM standards, the higher the certification level, and the higher the average reform score. Besides years of teaching, the only variable that reform score was significantly correlated with measured the percent of the school on free or reduced lunch. This correlation was negative, indicating that the higher the percentage of the school on free or reduced lunch, the less reform instruction occurring in surveyed classrooms in those schools. However, this correlation was fairly weak at only $-.356$.

Instructional Models

One of the intentions of this study was to characterize the instructional models used in different eighth-grade mathematics courses. To that end, teachers responded to a list of instructional strategies by identifying the approximate frequency with which they employ those strategies and the number of minutes they employ them during those times. Each strategy was classified as either a reform or traditional instructional strategy, in order to compose a number that would be representative of the percentage of class time that reform strategies were used. The 14 reform and 10 traditional strategies are listed in Table 14 for reference.

Table 14
Reform and Traditional Teaching Strategies

Reform

1. (B) Work on problems that have more than one solution.
2. (C) Make conjectures and discuss various methods during problem solving.
3. (F) Work together in pairs or small groups on mathematical problems.
4. (G) Work on group investigations that might extend over several days.
5. (H) Write about how to solve a problem in an assignment.
6. (I) Orally explain how to solve a problem.
7. (J) Work on mathematical problems embedded in a realistic context.
8. (M) Use manipulative materials or models.
9. (O) Use calculators or computers to solve problems requiring the integration of several concepts or skills.
10. (Q) Engage in student- or teacher-led whole group discussion.
11. (S) Complete any kind of **non-routine** items on tests.
12. (U) Complete items on tests requiring open-ended responses (e.g., descriptions, justifications of solutions).
13. (V) Engage in performance tasks for assessment purposes.
14. (X) Use calculators or computers as a tool (e.g., spreadsheets) or to explore patterns.

Traditional

1. (A) Practice computational or algebraic manipulation skills.
2. (D) Work individually on mathematics problems from textbook/worksheet.
3. (E) Memorize facts or steps.
4. (K) Practice application/word problems very similar to textbook examples.
5. (L) Practice mathematical rules or procedures.
6. (N) Work on single or two-step word problems.
7. (P) Listen to teacher lectures.
8. (R) Complete short-answer items on tests (e.g., multiple choice, true/false, fill-in-the-blank).
9. (T) Complete items on tests requiring symbolic manipulation and procedures.
10. (W) Use calculators or computers to practice skills.

Note. The letters in parentheses correspond to the lettered item on the questionnaire.

Overall Reform Scores among Three Levels of Courses

An average reform score for each teacher was calculated based on his or her responses to the series of instructional strategies listed in Table 14, which was included on both parts of the questionnaire. The score represents a percentage of class time spent on reform strategies, as perceived and reported by the teachers. Table 15 provides the means and standard deviations of the average reform score by course. None of the mean scores were significantly different from one another. It is worth noting that although the largest mean percentage of reform time occurred in Algebra II, the standard deviation in that course was quite a bit larger than in the other two courses, indicating the wider range of use of reform in those classes. The largest standardized effect size occurred between Algebra II and Algebra I and was .65. The effect sizes for the other two comparisons were approximately half that.

Table 15

Percentage of Class Time Engaged in Reform Instructional Strategies.

| Course | <i>n</i> | Mean (%) | SD (%) |
|-------------|----------|----------|--------|
| Pre-Algebra | 21 | 47.2 | 9.0 |
| Algebra I | 16 | 43.6 | 9.0 |
| Algebra II | 14 | 51.0 | 16.0 |

Note. The means are a percentage of total class time reported. The denominator was calculated by summing the minutes of use of all 24 instructional strategies.

Individual Teaching Strategies

In order to provide a picture of what instructional strategies teachers did report using, a mean percentage of time that each strategy was used was calculated for each course level. This was accomplished in the same manner as was the overall reform score. That is, the total minutes of use of each strategy divided by the total number of minutes reported gave the percentage of time any particular strategy was reportedly used by a teacher. Since this question was addressed on both parts of the survey, administered four months apart, there were two individual strategy calculations for each teacher. For those teachers who only completed those items on one survey, their individual strategy times were doubled before overall means were found. The mean percentages of class time for each instructional strategy across teachers were arranged in descending order; this was done for each course separately.

The means among all courses ranged from just under 1% to 13%, with standard deviations ranging from 1% to a high of nearly 8%. Thus, overall, there was not a lot of variation in the uses of instructional strategies as reported by the teachers. Ordering the 24 strategies by the amount of time they were employed showed that for all three levels of courses, the reform strategies were well intermingled with the traditional strategies. There are several ways to view the data to try to discern any differences among the courses, with regard to preferred instructional styles. The discussion that follows attends to the mean amounts of reported class time use, but also takes into account the rankings of the means of those strategies.

Out of 24 total strategies listed, 14 were classified as reform, whereas 10 were traditional. When the ordered list of instructional strategies by percentages of time

reported in use was broken into the top 12 (50%) strategies and bottom 12 strategies, there was a slight difference among the numbers of strategies in each of the two categories that were employed across the course levels. Table 16 lists those numbers.

Table 16 shows that the *number* of different reform strategies employed more often is slightly greater for the more advanced courses. That is, whereas, Algebra II used seven reform strategies in the top 50% of the time slots, Pre-Algebra used only five. For the bottom 50%, the case was reversed.

Table 16

Frequency of Reported Reform or Traditional Instructional Strategies by Class Time

| Course | Top 12 Strategies | | Bottom 12 Strategies | |
|---------------------------------|-------------------|-------------|----------------------|-------------|
| | Reform | Traditional | Reform | Traditional |
| Pre-Algebra (<i>n</i> = 42) | 5 | 7 | 9 | 3 |
| Algebra I (<i>n</i> = 32) | 6 | 6 | 8 | 4 |
| Algebra II (<i>n</i> = 28) | 7 | 5 | 7 | 5 |

Note: Overall, there were 14 reform and 10 traditional strategies listed.

A second way to examine the instructional models employed by the teachers is to examine and compare those strategies most- and least-often reportedly used in the different courses. Figure 3 shows the comparative amounts of time that seven different strategies were employed. These seven were chosen by their position in the top five most-employed strategies in any one of the three courses.

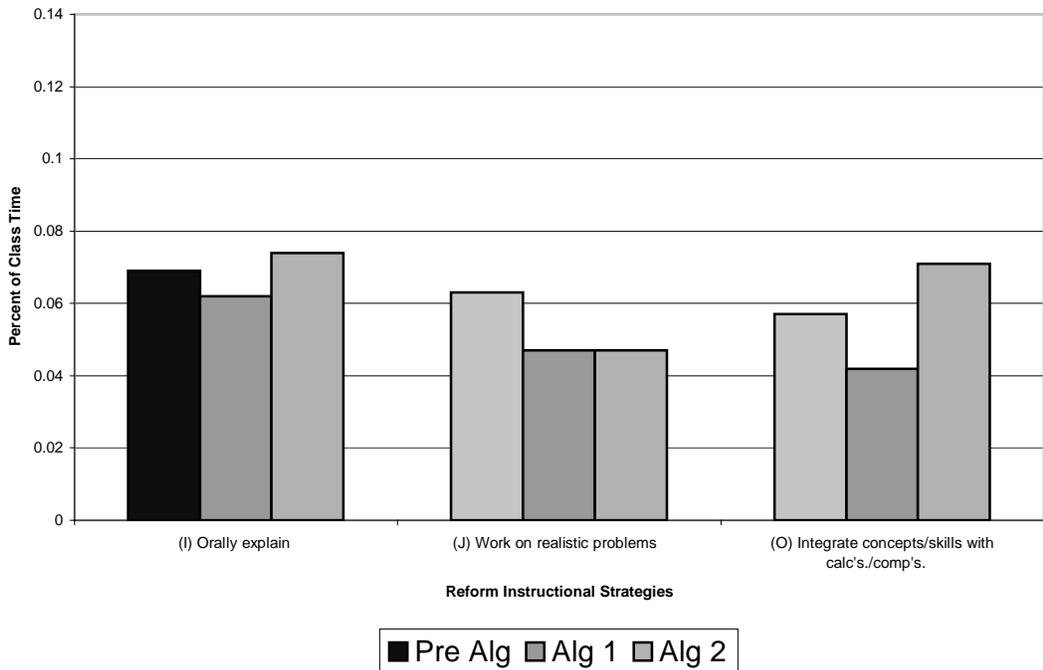
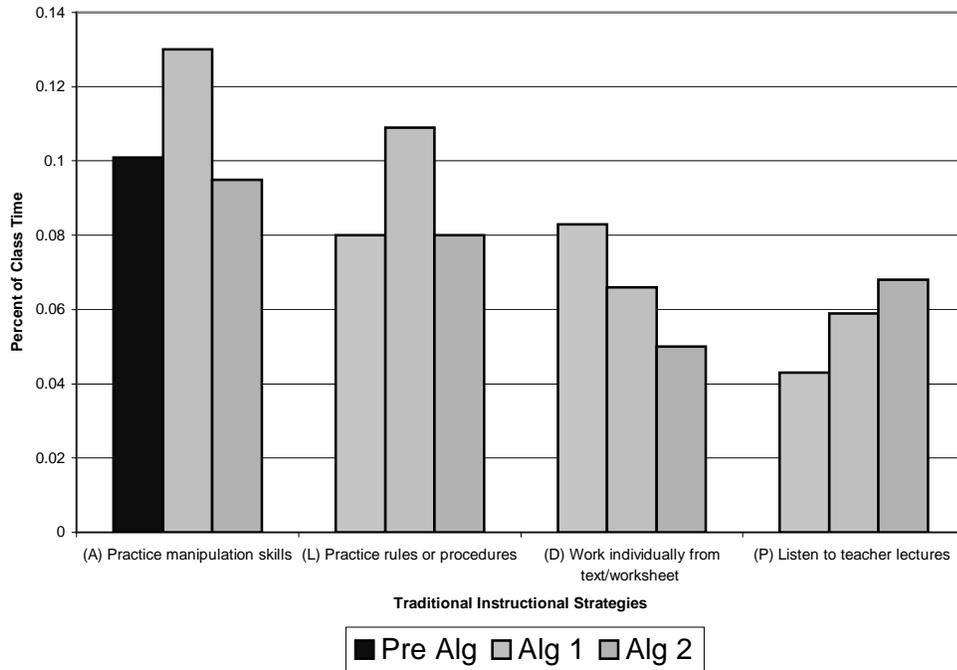


Figure 3. Traditional and reform instructional strategies that were among the reported top five most-used strategies in any one of the three courses.

The percent of time that Pre-Algebra teachers reported their students working individually on mathematics problems from a textbook or workbook, Item D, was approximately 3.3% more than for Algebra II students, which was a statistically significant difference at the .01 level. None of the other comparisons were significantly different.

Figure 4 shows the comparative amounts of time that the least-often used strategies were reportedly used. There were eight different strategies that comprised the bottom five in any particular course level. In this case, the amount of reported use of Item X, which was the practice of using calculators or computers as a tool or to explore patterns, was significantly greater for Algebra II classes than for the other two classes at the .01 level. In Algebra II, it was used approximately 3% and 2.3% of the time more than in Pre-Algebra and Algebra I, respectively. Although it was the third least-used strategy for the Pre-Algebra classes overall, which meant it ranked 21st out of 24 strategies, and the sixth least-used strategy in Algebra I, that item was ranked 11th out of 24 in the Algebra II classes. The only other strategy that was significantly different among any courses was Item H, which was writing about how to solve a problem in an assignment, but the difference was less than 1%. This strategy was also in the bottom 50% of strategies for all three courses.

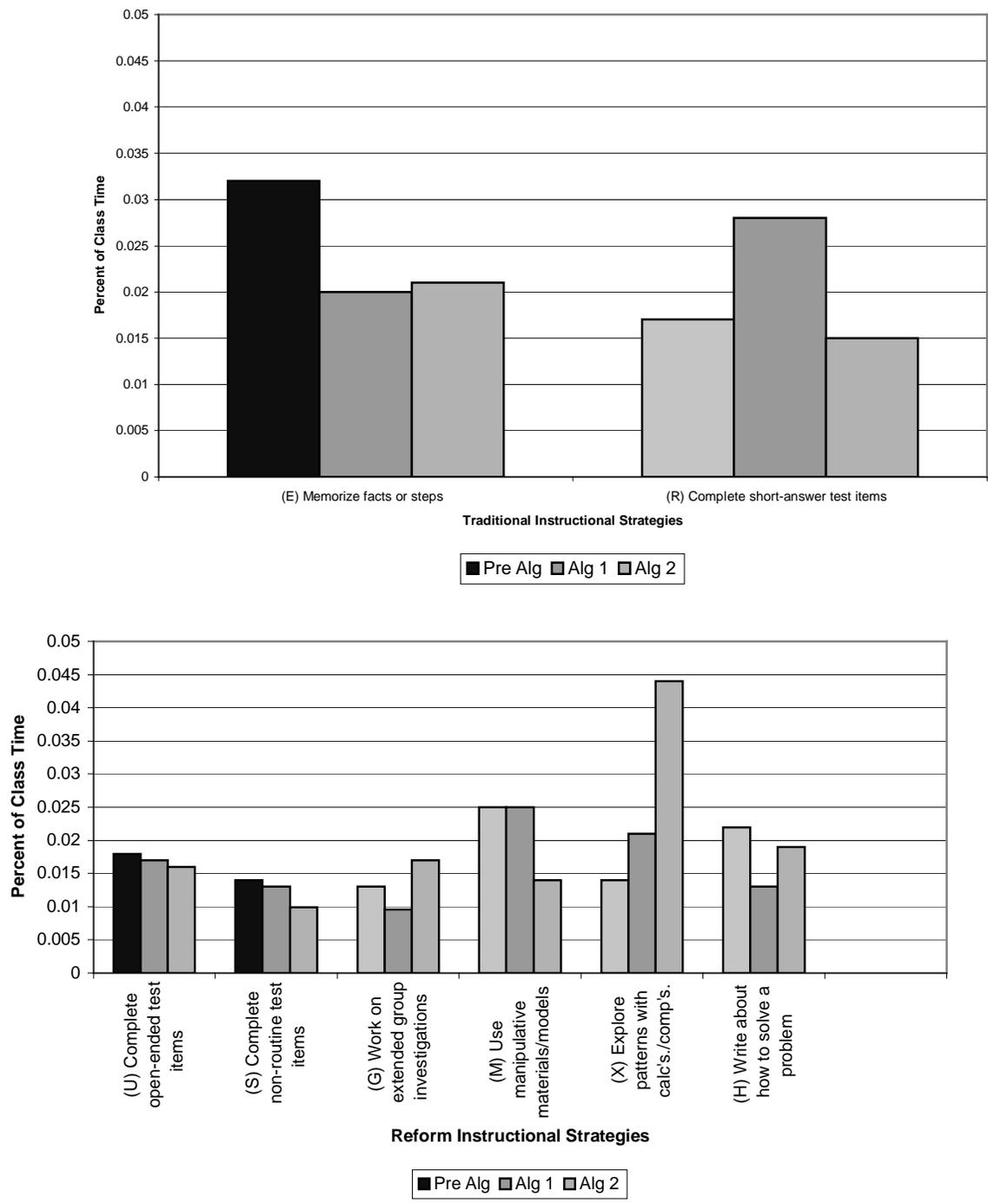


Figure 4. Traditional and reform instructional strategies that were among the reported five least-used strategies in any one of the three courses.

The three courses had more notable commonalities than differences.

Particular traditional practices dominated the top ranks of percent of class time reported for all three courses. Item A, practicing computational or algebraic manipulation skills, occupied the top time slot for all three classes. On average, it accounted for approximately 10%, 13%, and 9.5% of the class time in Pre-Algebra, Algebra I, and Algebra II respectively. Item L, practicing mathematical rules or procedures, was the second-most common instructional use of time in both algebra classes and was ranked third in Pre-Algebra. Working individually on mathematics problems from textbook or worksheet, Item D, occupied the second slot for Pre-Algebra and the third slot for Algebra I, but was further down the list in Algebra II.

As shown in the second graph in Figure 3, the most commonly used reform methods as reported by the teachers were Items I, J, and O. Those were respectively: Orally explain how to solve a problem; work on mathematical problems embedded in a realistic context; and use calculators or computers to solve problems requiring the integration of several concepts or skills. Item Q, engaging in student- or teacher-led whole group discussion, was the fourth most-used reform strategy. Those reform items that teachers reported as using more often might be considered the easiest reform methods to employ. For instance, more and more of teachers' resources, including textbooks are emphasizing "real-life" contexts, which was the crux of Item J. Also, it seems that teachers felt they were implementing the reform recommendation to require students to explain how they solved a problem, as implied in Item I. However, that reform strategy could be misaligned with the spirit of reform, depending on the focus of the teachers' questioning.

The least-used reform strategies and indeed the least-used strategies overall, as reported by the teachers, were those reform strategies shown in the second graph in Figure 4. Not surprisingly, these are some of the recommended reform strategies that would be hardest to implement without much professional development or on-going support. These include Item G, having students work on extended group investigations, and Item M, using manipulative materials or models. The other seemingly less-popular reform items dealt with non-traditional assessment techniques. These included Item S, having students complete non-routine test items; Item U, having students complete open-ended responses on tests; and Item V, having students engage in performance assessment. Teachers might have difficulty adapting their testing techniques to accommodate less traditional modes of assessment.

Predicting Reform Instruction

One of the main research questions for this study was, “What factors related to teacher, classroom or student characteristics influence the amount of reform instruction that teachers perceive they implement?” The analytic approach to answering this question was first, to complete a correlational analysis among variables in this study, including average reform score, and then subsequently to complete a stepwise regression procedure. However, as explained in the Methods chapter, a sample size of 51 limited the number of predictors that ought to have been tested in the analysis. Thus, the only variables that were entered as possible predictors were: years teaching grades 7-12 mathematics, percent of class on free/reduced lunch, PA/ teaching fit, CTBS mathematics score, and class problems. The only predictor to enter the regression

equation under the stepwise criteria of $p < .05$ for the F statistic was years teaching mathematics. The effect of years teaching was statistically significant, $F(1,47) = 13.30$, $p < .01$. The R^2 was .226, indicating that almost 23% of the variation in reform score could be explained by the teachers' years of experience teaching mathematics in grades 7-12. The unstandardized coefficient for years teaching mathematics was $B = .006$, which in practicality is quite small. That number indicates that for each year of experience teaching mathematics in grades 7-12, there is an expected increase of reform usage of .6 (just more than half) of a percent of total class time.

Achievement on the State Performance Assessment

Reform as Mediator or Moderator

The research question of interest in this section was whether, and in what way, the level of reform instructional approach influenced students' performance on the PA. One hypothesis put forth was that the level of reform instruction might have behaved as a mediator between the course level of the class and the achievement on the PA, since it was the case that class mean scores on the PA differed significantly from each other among the three courses at $p < .01$. Predictably, the mean scores were in the following order: 585.6 for Algebra II, 543.8 for Algebra I, and 497.0 for Pre-Algebra.

Nevertheless, in order for the reform-as-mediator model to be considered, a necessary pre-condition is that variation in the course level must account for variation in the amount of reported reform usage. In fact, based on the results from the ANOVA that showed no significant differences among the mean amounts of reform usage in the three courses, course level did not "predict" amount of reform instruction. That is, the

ANOVA amounted to a multiple regression analysis with reform as the dependent variable and dummy-coded courses as the independent variables. Thus, because course level did not predict reform instruction, reform could not be said to be mediating the relationship between course level and PA score.

Next, a regression of PA score on reform instruction was completed, first with all courses together and then separately for each course. When the entire sample, $N = 51$, was included, reform instruction did not predict PA score. However, in separate analyses, reform was found to be a significant predictor of PA score for Algebra II classes only ($n = 14$, $F = 6.7$, $p = .024$). In that case, the R^2 value was .363, indicating that the amount of reform instruction in Algebra II classes accounted for more than a third of the variation in PA scores. The effect size was such that a 10% increase in reform was associated with a 6.6 point increase in PA score.

The former result raised the question of whether a moderator model would describe the relationship between course level, reform instruction, and PA score. That is, was course level moderating the relationship between reform instruction and PA score, so that reform instruction predicted PA score, but differentially with respect to the specific course being taught? To determine this, a series of multiple regression analyses was completed in order to see if the interaction term of course level by reform instruction was significant after controlling for important background characteristics, such as sixth-grade CTBS scores and percent of class on free and reduced lunch.

Initially, a regression of PA score on course level only was completed using two dummy variables to represent the cases of Algebra II, or Algebra I, or neither of those, which would imply Pre-Algebra. The courses were significant predictors, $F = 62.471$,

$p < .01$, and the R^2 value was .722. Next, the regression was repeated with CTBS entering first, then course level. The purpose of this was to determine if course level remained a significant predictor once CTBS was in the equation. It was not. The CTBS scores alone accounted for nearly 93% of the variation in PA scores. Adding percent of class on free or reduced lunch added a significant change to the F value, with R^2 rising a small amount to .937.

Lastly, the moderator model was checked without controlling for background variables initially by entering in blocks, first the courses and reform instruction, and second, the two interaction terms (one for each course dummy variable). The interaction terms did not add a significant change in F over and above what was explained by course alone, and reform also was not significant.

Thus, the final model for predicting PA achievement was based on the two predictors, sixth-grade CTBS scores and percent of class on free or reduced lunch, shown in Table 17.

The result indicated in Table 17 is that nearly all of the variability in class achievement on the PA was explained by the sixth-grade CTBS scores. This points to the possibility that in fact CTBS score might be measuring the same thing as the PA. That notion is problematic, as the two assessments are given two years apart, and do not purport to measure the same aptitudes. In any case, CTBS score might not be an appropriate control variable in an analysis on achievement on the PA.

Considering that (a) CTBS score might be masking other predictors of achievement and (b) an earlier result pointed to an obvious disparity in teacher credentials, a post hoc analysis that examined whether teacher credentials was a

Table 17

Summary of Hierarchical Regression Analysis for Variables Predicting Achievement on PA (N = 51)

| Variable | <i>B</i> | <i>SE B</i> | β |
|---|----------|-------------|---------|
| (Constant) | -84.691 | 39.506 | |
| Step 1 | | | |
| CTBS Math | .957 | .057 | .869 |
| Step 2 | | | |
| proportion of class on free/reduced lunch | -33.480 | 13.159 | -.132 |

Note. R^2 for Step 1 = .929. $\Delta R^2 = .008$ ($p < .05$).

predictor for achievement was conducted. Because a linear model would not necessarily be appropriate for the ordinal nature of the credentials variable, an analysis was done using three dummy variables to code the four categories of credentials: provisional or emergency certification; non-mathematics secondary/middle certification, elementary, or special education; elementary education plus 18 credits of mathematics; and secondary/middle mathematics.

Course level was used as a control variable (in dummy-coded form) to determine if credentials added predictive value above and beyond the courses' effect on achievement. When certification level is the only variable in the model, $R^2 = .27$. Even when course level was included in the model as a control, teacher certification level was still a significant predictor of student achievement. Table 18 presents those results. The parameter estimates indicated that there was a rather large jump of 26 points from the lowest certification level (provisional or emergency) to the next highest. Another

Table 18

Teacher Credentials as a Significant Predictor for PA Achievement

| Variable | <i>B</i> | <i>SE B</i> | β |
|-------------------------------|----------|-------------|---------|
| (Constant) | 469.88 | 11.58 | |
| Step 1 | | | |
| Course Level | | | |
| GT status | 79.85 | 8.02 | .856 |
| Algebra I status | 39.461 | 7.26 | .45 |
| Step 2 | | | |
| Certification Level | | | |
| Mid/sec Mathematics | 37.81 | 12.15 | .41 |
| Elem + 18 math credits | 33.23 | 15.13 | .23 |
| Non-math or elem or spec. ed. | 26.19 | 14.06 | .21 |

Note. Provisional or emergency certification level was coded as 000 for this analysis, so it does not appear as a separate level.

R^2 for step 1 = .722. $\Delta R^2 = .05$ ($p < .05$)

analysis using only one certification coding of either certification above provisional/emergency or not, confirmed that there was a significant difference between those two categories. The parameter estimates for the certification levels also indicate the predictable outcome that the more mathematics a teacher had, the better their students fared on the PA.

Summary of Results

Differences Among Three Course Levels

Although, overall, only a relatively few significant differences were found among the three courses, the former results do illustrate a number of important patterns. First, Pre-Algebra teachers were clearly less credentialed in terms of their mathematics backgrounds than were algebra teachers. Of the 20 Pre-Algebra teachers that responded to that question, 40% of them did not have secondary/middle mathematics certification, nor had they earned at least 18 credits of mathematics coursework. That number compares with only 12.5% of Algebra I teachers, and no Algebra II teachers who were lacking that level of certification. In the end, an analysis showed that the teachers who were essentially uncertified tended to have students with lower PA scores, even when taking into account their course status.

The conflict variables revealed an interesting pattern, although not all differences found were statistically significant. Still, compared to the other two courses, Algebra I teachers had the lowest mean score on the accord factor (accord between teaching to prepare for the state PA and simultaneously covering their course content), and reported the lowest mean score on being able to engage their students in higher-order thinking. Those teachers also had the highest mean for mixed messages with regard to teaching their classes, while preparing their students for both the Algebra I final exam and the state PA. These teachers also reported the most diversity among their classes. It appears that Algebra I teachers, more than the teachers of the other courses, believed their course to be less compatible with expectations of the state PA in terms of instructional emphasis. They seemed to have a more difficult time juggling preparation

time for all of the end-of-year assessments their students would face. Perhaps the fact that they reported having more diversity in their classes also led them to feel more pressure as far as issues that compete for instructional time and strategies in the classroom. Accordingly, Algebra I teachers received the lowest reform score for reform teaching among the three courses, although not by a significant amount.

On the other hand, Pre-Algebra teachers reported the highest mean score, by a statistically significant amount, for the match between being able to prepare their students for the state PA and teaching the way they wanted to. Yet, these Pre-Algebra teachers also reported a significantly greater influence on their instruction by the FMT. Their score for reform teaching was almost exactly halfway between Algebra I teachers and Algebra II teachers, though none of the differences were significant. In any case, several of the results seemed to contradict one another.

Instruction in Arbor District Middle Schools

Taken as a whole, variation among reported use of various instructional styles was modest. Each individual traditional or reform strategy was reportedly being used for anywhere from 1% of class time to a high of 13% of class time. However, traditional teaching strategies, such as practicing computation and manipulation skills and practicing rules or procedures, were reported to be the most commonly used strategies by the teachers.

Of the five assessment strategies listed, performance tasks were the most commonly used strategy in Pre-Algebra. Algebra I and Algebra II reported more traditional strategies as their most common assessment techniques: short answer items,

and items requiring symbolic manipulation. However, those results speak to the rankings of the strategies; there were no significant differences between the mean amounts of class time that various assessment strategies were used across the courses.

Instructional Styles and State Assessments

Interestingly, the only significant predictor of the amount of reform instruction was years teaching mathematics in grades 7-12. The slope was positive, so that the more years of experience, the higher percentage of class time was reportedly spent using reform strategies. Practically speaking, however, the amount was very small at only just over a half of a percent of reform time for each additional year of teaching. This result is somewhat surprising, in that there are other factors related to years of experience that would seem to be more directly related to implementing reform strategies, such as professional development or recent graduate work in mathematics education. Yet those variables were not correlated with reform. Also, the opposite trend might have been hypothesized – that younger teachers who have gone through more recent teacher education programs would be more inclined to use reform strategies, in accordance with the current trend in mathematics education. Although the relationship was only modest, with $R = .475$ and $R^2 = .23$, it remains curious and an issue for further research.

There was a clear relationship between course level and achievement on the state PA. At the beginning of this study, it was hypothesized that specific instructional strategies might mediate the relationship between course level and achievement. However, this study did not reveal that to be the case. When PA achievement was regressed on reform instruction separately for each course, reform instruction was found

to be a significant predictor for Algebra II only. However, when interaction terms were used with all courses together to examine whether a moderator relationship was at work, no significant difference was found. This indicated that although the slope for reform instruction was significant and positive for Algebra II (in Algebra II, the more reform instruction, the higher the achievement on the PA), the slopes for the three courses were not different enough from one another to be found significant, or there simply was not enough statistical power to provide evidence that a real difference existed. Thus, instructional style acted neither as a mediator nor as a moderator influencing the relationship between course level and achievement on the PA. In fact, the sixth-grade CTBS scores was an immensely strong, and nearly the only, relevant predictor of achievement ($R^2 = .93$). That result is disappointing in the sense that between the end of the sixth grade and the end of the eighth grade, the achievement gap appeared to be unaltered. The mathematical content assessed on the state PA is material to which all eighth-graders would have already been exposed at some point in the students' academic careers. Many contend that the test has rigorous expectations of students with respect to analysis, synthesis, and communication of mathematical problems. It is rather discouraging that a sixth-grade skills test should be so predictive of an eighth-grade performance assessment.

CHAPTER 5: CLASSROOM OBSERVATIONS

Method

The purpose of the observational component of this study was to document the kinds of things that exemplary eighth-grade mathematics teachers are doing in the classroom. The main research questions for this component of the study are: (a) What are the characteristics of instruction in the classrooms of eighth-grade mathematics teachers identified as “exemplary” traditional and reform teachers? (b) Are there any comparisons or contrasts in the high-track and the low-track classrooms with respect to teaching for conceptual understanding, particularly for two levels with the same teacher? (c) Are the instructional activities and discourse aligned with the expectations evident on the state performance assessment? That is, do students in the classrooms observed appear to have the opportunity to learn the skills assessed by the mathematics component of the PA?

The goal was to observe teachers who were described as having either traditional teaching styles or reform styles and to observe them teaching two different levels of eighth-grade mathematics classes. Thus, names of teachers were originally solicited from principals through a letter (Appendix D) asking them if they could recommend eighth-grade mathematics teachers whom they consider to be exemplary. During a telephone conversation, principals were asked to express their reasons for identifying a particular teacher, without any attempt to direct their answers toward any particular teaching style. Their responses were recorded nearly verbatim.

To address the issue of what teachers do in their classrooms, teachers were observed teaching two different classes for an inclusive period of three to four weeks. In

most cases, they were observed on consecutive instructional days, omitting days when a full-period test was administered. Originally, my intention was to space the observations so that I could observe the teachers introducing a unit or topic, developing the concepts embedded in the unit or topic, and bringing it to closure. However, because the teachers were in different schools all over a county, I could only observe one teacher at a time, which did not always coincide with the beginning of a unit. I did, however, observe each of them over a sufficient time frame to permit the opportunity to see all of the teachers introduce something new to the students, develop concepts (or fail to do so), and end the teaching of a topic with or without closure.

The data collected consisted of extensive field notes taken during classroom observations using a specific observation protocol for all observations. Observation forms (Appendix E) identified categories that segmented the class time and characterized a variety of lesson components. Field notes consisted of specific dialogue and teacher and student actions and behavior, as well as any immediate comments about those observations. These observations and comments were recorded using a two-column format. Post-observation notes (Appendix F) were used to reflect on field notes taken during observations and to focus on the aspects of the class that were proposed to be under the most scrutiny: student tasks, discourse, and learning environment. Often, typed narrative notes were composed immediately following an observation in order to summarize both observations and reflections. The questions that guided the interpretation of observations were “What opportunities existed for students to develop conceptual understanding in mathematics?” and “What kinds of expectations does the teacher seem to have for his or her students?” Additionally, all instructional materials

handed out by the teachers were collected. Finally, informal conversations with teachers occurred on an on-going basis before or after instructional observations, and these were written up as additional field notes as soon as possible after taking place.

The framework that was used to analyze the classroom observation data is described in detail in the data analysis section that follows a brief description of the sampled teachers. This framework is consistent with over a decade of research and theory, most notably spirited by the National Council of Teachers of Mathematics (NCTM). That is, the goal of instruction is to teach for understanding; students ought to be developing conceptual understanding, as well as procedural knowledge (Hiebert & Carpenter, 1992). Research in education and cognitive psychology indicates that learning for understanding promotes flexible thinking, fosters the construction of relationships among concepts, and enhances transfer. Research also suggests the conditions that foster learning for understanding. In particular, learning should frequently be an active process in which students engage in problem solving; learning is greatly enhanced through interaction; new knowledge must be connected (by the learner) to existing knowledge; and isolated pieces of information must be connected into coherent structures (Brooks & Brooks, 1993).

Instructional practices are likely related to teacher cognitions and beliefs concerning content and pedagogy, as well as how students learn (Artzt & Armour-Thomas, 1999). The present observational study is just that – observational. Teacher *behavior* was studied through my observations as a single researcher. Very limited data was collected on teacher *cognition* through informal conversations that I had with the teachers. Furthermore, I could not formally assess what the students actually learned in

the classes in which I observed; my only glimpse into what students understood was gleaned from careful attention to observed interaction of students with the teacher or other students. Thus, this study only sought to describe in detail what kind of instructional practices occurred and could merely *conjecture* about the reasons for certain teacher behavior and the likely implications for student learning. However, these conjectures were based on verbalizations, either by a teacher or students.

These observations were further limited by my own ability to observe and attend, which was different in different classrooms. For instance, in one of the classrooms, due to the arrangement of the room and where I was sitting, I could only see the face of one student and generally could not clearly hear conversations between students or “private” interactions between selected students and the teacher.

Despite these limitations, however, there are advantages of observations over other methods of assessing instructional practice, such as interviewing and surveying (Creswell, 1994). I did not need to depend on the teachers’ *perceptions* of how they teach. Rather, I was able to gather firsthand experience in the classrooms of these teachers and could record information as it occurred. Whether the teachers were comfortable with mathematical content matter or not, I was able to observe how they attempted to “transfer” knowledge to students. I was also able to notice and record many student reactions or responses that were not acknowledged and may not have been recognized by the teachers.

Subjects

The observational subjects of this study were three middle school mathematics teachers in Arbor County School District-- Ms. Drake, Ms. Miller, and Ms. Henderson (pseudonyms) -- who had all answered the teacher survey. Ms. Drake and Ms. Miller were identified by their principals, and Ms. Henderson was identified by the mathematics specialist in the district Office of Mathematics.

Principals in the 20 schools participating in the quantitative study were sent a letter in early January soliciting the names of eighth-grade mathematics teachers whom they considered to be exemplary. Only three principals responded, one giving the names of two teachers in his school. Judging by the comments made by the three principals, the two teachers from the same school were characterized by approaches that were more traditional in nature, whereas the other two teachers were described as having styles that, at least nominally, were more commensurate with reform ideals. The goal was to observe one teacher that tended toward traditional approaches and one that exemplified more reform-oriented instruction. A letter of inquiry was sent to all four teachers, soliciting their permission for classroom observations. However, the intent was to observe only two teachers.

I began observing Ms. Drake, who was identified as an exemplary teacher by her principal, in the second week of February. After about one week of observations, I began to question whether principals who did not necessarily have mathematics backgrounds could accurately evaluate the pedagogical content knowledge of mathematics teachers. I could not solicit the input of school-based mathematics content leaders (similar to department chairs) because they themselves generally taught eighth-

grade mathematics, and I did not want to put them in the position of having to identify themselves or purposely to not identify themselves.

While I continued to observe Ms. Drake, I met with Ms. Byrd, the specialist in the district Office of Mathematics to solicit her recommendations for exemplary mathematics teachers. I explained to her that I was interested in observing exemplary teachers who would be considered either traditional or reform in nature. As a mathematics educator, she understood my intention and was able to identify teachers in both categories, explaining her reasoning. As a caveat, Ms. Byrd first explained to me that I would likely notice a certain traditional-looking structure to *all* classes across the district, no matter what subject was taught. She said that the area and school expectations had been stressing for many years a lesson structure that followed to some degree the following: drill/warm-up, teacher-directed instruction, guided practice, closure, and assessment. She told me explicitly that this was not a directive from the Office of Mathematics, and that their office was hoping to have in-service workshops to help teachers adjust this structure to allow for more classroom interaction and for reform practices in general.

Ms. Byrd identified two teachers from one school, both of whom she was considering “traditional,” although she said that in actuality they were excellent at incorporating the best of the reform models, while sticking to a traditional format of teaching. She suggested I contact them myself, using her as a reference, and that more than likely one of them would be willing to participate. She also identified two teachers whom she thought were “the most reform” in style in the district, as well as very knowledgeable in their content area and effective teachers. She said that she would

contact them for me, as they might be more hesitant about participating. The result of that meeting was that Ms. Henderson agreed to be observed. Ms. Henderson was one of the “traditional” teachers that I contacted. Both of the teachers that were considered to be effective reform teachers refused participation. Therefore, I would need to observe one of the two teachers originally identified by the principals, neither of whom were particularly identified as having a strong reform style by Ms. Byrd. It was not that Ms. Byrd disagreed with the principals’ assessments; rather she was unfamiliar with the teaching of these individuals.

Of the two “reform” teachers identified by the principals, one was very apprehensive about participation due to “feeling generally overwhelmed,” and one readily consented. I explained to all of the teachers with whom I spoke that I would be observing their class, taking notes, and only occasionally asking them questions for clarification after class. They would not be asked to do anything special with regard to this [observational] study. After Ms. Drake declined my request to audiotape her lessons, other teachers were told that this would not be requested. The “reform” teacher that agreed to my observations was Ms. Miller.

Although my original intention was to observe only two teachers, I will discuss my observations of all three teachers because all three manifested different teaching characteristics.

Ms. Drake

Ms. Drake was in her late forties and had been teaching for 24 years, with nearly all of her experience in the middle school in which I observed her. She was then the

mathematics content leader (similar to a department chair) for her school. Her principal called me to specifically recommend her. He noted that she was teaching the upper level classes in the school and was “very straight forward.” He stated that she had a “traditional style” and “stayed on top of kids until they got it.” He explained that she “modeled thinking” for her students. He also said that she had great success with kids on the FMT, and that the school in general was not successful on the state PA. On the first day of my observations, an office secretary walked me to Ms. Drake’s classroom and, as she did, she mentioned how lucky I was to be observing Ms. Drake because she was really the “cream.” The secretary told me that Ms. Drake was able to get students, whom others considered hopeless, to pass the FMT.

Ms. Henderson

Ms. Henderson was likely in her mid-forties, and had been teaching middle school mathematics for 10 years. The specialist in the district Office of Mathematics recommended both Ms. Henderson and her mathematics content leader, calling them “peas in a pod” – both of whom she considered to be exemplary and similar in style. At first, the district specialist described these two teachers as traditional, but reconsidered that label, and said that they kept the best of “traditional” teaching, while balancing it with many reform ideals. She further described them as life-long learners, always seeking professional development beyond what is required by the district. When I spoke with the content leader at the school, she suggested I observe Ms. Henderson, mainly because Ms. Henderson was teaching the most advanced eighth-grade course (Algebra II), as well as the lowest level eighth-grade course (Pre-Algebra). She also said that she

had learned much of what she knew from Ms. Henderson, and that they planned together for Algebra II. She also mentioned that they very often had observers and student teachers in their classrooms.

Ms. Miller

Ms. Miller was in her mid to late twenties and had been teaching at this middle school for 5 years. Her principal contacted me in response to the letter I sent out soliciting names of “exemplary” teachers. He told me that he thought she was “excellent.” He went on to say that she worked well with all types of students and described her as follows: She gave students challenging problems that went beyond what they had been asked to do previously; she didn’t let students give up easily; she answered questions with questions; she gave plenty of word problems that required students to read and break down the problems. He said that she asked students to write their own notes, detailing their steps in problem solving, rather than telling the students the steps herself. He said that she was “strict.” She also gave activities that were similar to ones that were on the state performance assessment. He had observed her doing an exercise with students in which she had them write a letter to an imaginary student who had missed class, so that they needed to explain what they did or learned during class. He said, “Students like her. She respects students.”

Ms. Miller readily agreed to my requests to observe her classes. She taught only Algebra I and Algebra II, and I would be observing one particular section of each course.

Data Analysis

The classroom data consisted of field notes taken during instruction, reflections composed directly after observations, and instructional handouts from teachers. Since each teacher was observed teaching two different courses, their data were sorted by the course observed. For each observed course, the data were analyzed according to the following categories: tasks or activities, learning environment, and patterns of discourse. Those categories are nearly identical to the “lesson dimensions” defined in a study by Artzt and Armour-Thomas (1999) in which a model was developed to examine mathematics instruction. These authors further broke each dimension into indicators.

In the Artzt and Armour-Thomas (1999) study, for the dimension entitled “tasks,” indicators were (a) modes of representation, (b) motivational strategies, and (c) sequencing/difficulty level. Broadly speaking, this category considers whether the tasks posed by teachers are at the appropriate level for students to take into account what they already know and to extend their knowledge; this category also notes whether the tasks incorporate multiple representations or modes of analysis. For “learning environment,” the Artzt and Armour-Thomas indicators were (a) social/intellectual climate, (b) modes of instruction/pacing, and (c) administrative routines. That is, the climate of the classroom ought to be a place whereby students and teacher respect each other’s opinions and teachers foster and encourage students to explore mathematical ideas. A variety of instructional strategies as well as effective classroom management would support this goal as well. For “discourse,” indicators were (a) teacher-student interaction, (b) student-student interaction, and (c) questioning. To fare well along these indicators, a teacher needs to listen carefully to students’ evolving conceptions and

address them appropriately; teachers must require that students justify, explain, or clarify their responses orally or in writing and must make appropriate decisions about when to give more information to students or when to allow students to grapple with ideas. Additionally, teachers must encourage students to interact likewise with other students.

Often in this study, other themes regarding the teaching of mathematics emerged for the various teachers and classes. For instance, for two of the teachers, level of content knowledge as well as attitude toward students and content became apparent as issues affecting their instruction.

In the following sections, I address each teacher separately, at first separating the two courses observed. Teachers' practices will be described in detail so that the reader has a clear picture of what was observed. When quotations are used, they are enclosing an exact or near exact verbatim account of the dialogue that took place. (Although audiotapes were not used, I was able to write much of the dialogue as it occurred.) Following the description and commentary for each of the classes, the teachers' overall instructional practice will be considered according to the categories named above. This review will serve to address the main research question: (a) What are the characteristics of instruction in the classrooms of eighth-grade mathematics teachers identified as "exemplary" traditional and reform teachers? For each teacher, I will also address two of the main research questions: (b) Were there any comparisons or contrasts in the high-track and low-track classrooms with respect to teaching for conceptual understanding, particularly for two levels with the same teacher? and (c) Were the instructional activities and discourse aligned with the expectations evident on

the state performance assessment? That is, did students in the classrooms observed appear to have the opportunity to learn the skills assessed by the mathematics component of the PA?

After I have addressed each of the teachers in isolation, I will discuss all three qualitative research questions, comparing and contrasting the three teachers.

Results

Ms. Drake

Ms. Drake was teaching in a middle school registering about 600 students. The school had a minority enrollment of 34.5 %, and just over half of the students received free or reduced lunch, with the vast majority being free. The school had the third largest percentage of students on free or reduced lunches in the district. I observed Ms. Drake teaching a course called “Fundamentals” (pseudonym) and an Algebra II course. These were both the lowest and highest level courses in the eighth grade. In between the two classes that I observed, Ms. Drake had a two-period break. With her consent, I generally stayed in the room and did paperwork, while she came in and out, taking care of school or personal matters. During this time, however, I often was able to talk to her about the classes or students. Very often she would initiate conversations that tended to center on students’ deficiencies or personal problems, as opposed to concerns related directly to teaching mathematics. When I refer to comments she made to me or short conversations we had, they generally occurred during this break time.

Ms. Drake’s room was arranged with three columns *of pairs of desks* facing the front of the room. That is, desks were arranged two-by two, butting side by side, facing

the front of the room so that three columns were formed. Because of the size of the Fundamentals class, this meant that most students had someone sitting directly next to them. In the Algebra II class, since there were only 14 students, each student was generally seated at his or her own pair of desks. I do not know who arranged where students would sit.

Fundamentals

I visited Ms. Drake's "Fundamentals" class for a total of seven observations over a three-week period. The course was the lowest level of eighth-grade mathematics content offered at that school, enrolling students who were not placed in Algebra I. The course content was a mixture of geometry, statistics and probability, and pre-algebra. The particular class I observed had many students who still had not passed the FMT, whereas the other fundamentals classes had students who, by October, had all passed. (The mathematics department actually re-scheduled students so that mathematics Fundamentals classes were more homogeneous after the October administration of the FMT.) Thus, I was observing the students considered to be the lowest performing eighth graders in mathematics. There were 23 students in the class. Table 19 shows the three-week span over which I observed, and whether an observation took place or the reason one did not.

Ms. Drake's Fundamentals class followed a general structure that segmented the 45 minute period. As students walked into the room, there was a "drill" exercise on the overhead projector that generally took about five or six minutes. Following this, they usually read homework answers aloud consuming on average five more minutes of class

time. Generally, two to five minutes were then spent on other administrative procedures,

Table 19
Time Table for Observations of Ms. Drake’s Fundamentals Class

| Date | Observation / No observation |
|------------------------|--|
| Thursday, February 10 | Observation |
| Friday, February 11 | No observation – teacher’s suggestion. Ms. Drake said students would be working out of the book to practice using calculators with order of operations. |
| Monday, February 14 | Observation |
| Tuesday, February 15 | Observation |
| Wednesday, February 16 | No observation – teacher’s suggestion. Ms. Drake said the period would be taken up with a quiz. (In fact, they did not get to the quiz, since she felt they needed “more practice.”) |
| Thursday, February 17 | Observation |
| Friday, February 18 | No observation – Snow day |
| Monday, February 21 | No Observation – President’s day |
| Tuesday, February 22 | Observation |
| Wednesday, February 23 | Observation |
| Thursday, February 24 | Observation |
| Friday, February 25 | No observation – teacher’s suggestion. Ms. Drake thought most students would be away on a field trip. |

such as giving quiz answers, handing out materials, or copying headings onto their “table of contents” sheet, used by students to keep their notebooks in order. (Notebooks were periodically collected so that Ms. Drake could make sure all required homework,

classwork, quizzes, etc. were there.) Of the seven days I observed, students took a quiz on three of the days. The quizzes took 15-17 minutes at the end of the class period. On two of the quiz days students completed a review sheet, which was very similar to the quiz, immediately preceding the quiz. The reviews took about seven minutes for individual work and then seven minutes for the whole class to go over. On the four days when there was no quiz, the rest of the period generally was broken into a whole class lesson that lasted from 8 to 21 minutes, followed by individual seatwork that lasted from 5 to 21 minutes. On two of the days with lessons, I observed a short “wrap-up” that lasted about three minutes.

In the Fundamentals class, initial drill exercises typically consisted of two multiple-choice problems dealing with fractions, decimals or percents. This was meant to give students practice with skills that are tested on the FMT, as evidenced by several comments Ms. Drake made after going over the answers. (“They always give the same [answer] with a decimal! Don’t pick it!” Another time she said, “Remember what I told you back in August and September. $\frac{1}{3}$ and $\frac{2}{3}$ are the ones you have to memorize since one or the other will appear on the test.”) The drill segment generally took about 5 minutes. On some days students put the answers on the board; other times Ms. Drake worked the problems on the board and still other times, only answers were given. She always asked if anyone had any questions, but no one ever asked any. Ms. Drake did not require students to show all their work on the board, nor did she ever ask students to explain what they did. There was rarely any discussion at all about the drill exercises, other than Ms. Drake quickly mentioning a procedure or mnemonic device out loud.

On several occasions, she referred to money as a way of explaining the problem, but she never showed other methods or solicited any from students.

Homework was covered in a similar fashion to that of the drill, except that students switched papers with someone nearby and marked items right or wrong. Ms. Drake either read the answers aloud or called on students who had their hands raised to do so. If wrong answers were read aloud, another student was called on for the correct answer. Wrong answers were never discussed. Of the seven days I observed homework being corrected, the only two questions ever asked came from the same student on the same day. Both questions asked about procedure; Ms. Drake responded by reiterating the procedures in question.

On four of the seven observations, there were lesson segments in which either new concepts were taught, or new kinds of problems were introduced. The first lesson segment I observed was an introduction to the unit on algebra, in which a handout entitled “Language of Algebra” was used for the whole-class lesson. It focused on definitions and recognition of equations, expressions and inequalities. For example, typical problems on the handout were “Tell whether each is a numerical expression, an algebraic expression, an equation, or an inequality: $2 < 5 + 9$;” and “Name the variable in: $2y + 3 = 8$.” The three subsequent lessons expanded on the ideas in that first lesson by giving examples of different kinds of problems that were expressions, equations or inequalities. All of the lessons were followed by seatwork wherein students worked through problems in the text individually, stopping at the end of each section of problems until Ms. Drake told them which of the next few problems to try.

Overall, the lessons were characterized by a very teacher-directed instructional style. Furthermore, those lessons lacked cohesiveness and failed to develop concepts based on sound mathematical notions. In some cases, students were told to read a small portion of the book; Ms. Drake would then go over the material that was read, with little if any elicitation of students' ideas; mathematical definitions she gave were either inadequate or inaccurate; she failed to relate new ideas to prior knowledge or realistic contexts; she used manipulatives inappropriately and even incorrectly; she provided students with little more than trivial "do as I do" practice problems; and there was very little interaction between her and her students. The few questions that Ms. Drake asked were merely seeking answers to the practice problems that either she or the book posed. In general her explanations seemed inadequate. I will describe the first new lesson of the algebra series in its entirety and highlight some revealing segments during the other lessons in order to illustrate some of these weaknesses. My comments are in parentheses.

In the first lesson, "language of algebra," students were told to turn to a page in the book and to read a specific section. She told them, "You are reading to be informed. Do not write." She gave them a few minutes to read when she noticed that one student was not reading. She went over to him and directed him to read, then walked over to me and quietly said, "The laziness wears on me after awhile." She spent about one minute talking about how to read a mathematics book, and asked why some words were in bold. One student said, "important terms." She added, "They give examples." She then went over what they read – definitions and examples of algebraic terms. She wrote them on the overhead, and asked students to do the same on their handout. Her definition of a

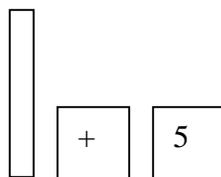
numerical expression was “an expression that only contains numbers.” She solicited examples from students. Although operations were not mentioned, one student suggested “ 20×18 ” and another offered “ $5 + 6$.” She went on to define a variable as “a letter that represents a number.” Again, she asked for an example and two students named the letters “a” and “d.” She said, “Any letter you can think of.” She did not bring up the idea of a variable representing an unknown quantity, or the notion that a variable is useful when a quantity can change from one instance to another. Next she defined algebraic expression as “an expression that is written with a variable.” When she asked for an example, a student said “a 3 plus 5.” She corrected him, saying to put the number first, and then she wrote “ $3a + 5$ ” on the overhead. (There was no discussion of the operations, or what “a 3,” or for that matter “ $3a$ ” might mean. It is not likely that the students interpreted “ $3a$ ” as implied multiplication, since this was their introductory lesson to algebra.) Her definition of equation was “a statement that uses an equal sign.” Two students gave the examples of “ $6 + 2 = 8$ ” and “ $90 - 4 = 86$.” Ms. Drake wrote those down and added “ $a + 7 = 10$ ” with no explanation. She then went on to define the five inequality symbols ($>$, $<$, \leq , \geq , and \neq).

The students had a handout that presented vocabulary and two types of problems in two sections. The first section listed three algebraic equations or inequalities. Ms. Drake asked students to name the variable in each. They did so, by merely identifying the letter they saw. There was no discussion of the meaning of any other mathematical symbols in the mathematical statements, nor any attempt to connect the meaning of symbols. The second group of five problems asked students to tell whether each given item was an equation, expression or inequality. They did this aloud, and the reasons

given were based solely on the symbols (+, <, none) that they saw. Next, Ms. Drake told them, “Take out a sheet of paper. Number your paper 10 to 25; don’t skip lines.” She asked someone to read directions, after which they began working individually on the problems that were identical in kind to the ones they had just completed aloud. They worked individually as Ms. Drake went around the classroom to check their work. The focus of the lesson was on recognizing symbols, as I heard her say to a student, “Does this have an inequality? Then it can’t be that, right?” To another student she asked, “What do you notice this has? Then it must be...” After five minutes of working individually, either she or students called out answers with no questions or discussion. She told them their homework was to complete these problems. The class ended.

The next day’s lesson addressed the same material, showing students how to do various kinds of problems that involved the same symbols and definitions they had already learned. However, some incidents during this lesson are worth describing, as they illustrated sequencing decisions and potential opportunities for developing conceptual understanding.

Ms. Drake began the lesson by asking for an algebraic expression. No one responded. She used algebra tiles to put the following display on the overhead:



She named it “ $a + 5$.” A student called out, “ a ?” to which she responded, “I can use any letter I want for a variable.” (Ms. Drake did not probe to determine what the student was asking by “ a ?” Ms. Drake introduced the algebra tiles in the *next* lesson on algebra.)

This whole-class portion of the lesson lasted eight minutes. The remainder of the period was spent on bookwork with more problems related to equality and inequality symbols. For each series of questions, she did the first example of the kind and then let students work individually to complete that section. One group of problems asked students to turn expressions into equations. (If the expression contained only numbers, students had to write “=” followed by the correct number. If the expression contained a variable, they could write “=” followed by any number or expression.) The students seemed confused. Ms. Drake then went to the overhead, wrote “ $x + 4$ ” and asked the class, “How do I make it an equation?” A student responded, “Pick a number for x .” She ignored his comment, and said, “We need ‘=,’ right?” Then she told them they could put any number on the right side. She asked, “Why doesn’t it matter what number I put here?” She answered her question by saying, “Since x represents a number and we don’t know what it represents!” There was no further discussion or explanation of her statements.

This interchange was the first time she talked about a variable representing an unknown; in the previous lesson she stressed that a variable was a letter. As Ms. Drake monitored individual seat work, I could hear students asking her questions which indicated confusion about the concept of a variable. The drill problems they were completing did not address this confusion. Furthermore, there was no discussion of what students had created once they “completed the equation.” That is, Ms. Drake did not lead a discussion or ask questions to focus students’ attention on the fact that whatever number they chose for the right side of the equation changed the value of the variable – an important and fundamental concept.

The topic of the third lesson was translating between English and mathematical expressions. For example, the English expression “a number increased by seventeen” could be written mathematically as “ $a + 17$.” In this lesson Ms. Drake also introduced algebra tiles.

Ms. Drake introduced the third lesson by saying “So far in algebra, you have identified equations, inequalities, and expressions. Today we are going to learn to write expressions.” After eliciting synonyms for addition and subtraction by giving hints such as “starts with...,” she presented the algebra tiles. She displayed a long rectangle and told the students that the green side represented a variable. She told them that “the little yellow square means addition,” and “the little red square means subtraction.” (She did not relate the algebra tiles to an area model, which ultimately gives the tiles concrete meaning, and misnamed the “little squares” as operations, rather than as positive and negative units.) She then said, “Let’s model $x + 3$.” She immediately showed them a long rectangle and three small yellow squares. The room was silent. (According to her definitions, a student might have thought that her model showed “ $x + + +$,” which would seem meaningless.) She then presented a series of problems in which she wrote an English phrase and asked students to show it in tiles and then to write it with symbols. She reminded the students of the commutative property when she asked them if it was alright to put the little squares in front of or behind the rectangle for the expression “3 more than a number x .” (That is $3 + x = x + 3$.) However, she told students that for “3 less than x ” the tile representation cannot be switched. (This was a mistake. Although $(3 - x)$ and $(x - 3)$ are different, the tile representation for $x - 3$ is the

same whether the small chips are in front or in back because the chips actually represent negative 3 no matter where they are.)

There were no other questions asked of her or the students during this lesson. The only interaction that took place was at the end of class when she tried to assess what they knew. She put on the board “ $P + 9$ ” and tried to elicit English phrases. Someone said “ P is greater than 9.” She said, “No” without further explanation. The students did not respond to her requests for other phrases. Instead, she named aloud four different phrases that mimicked the original key words given at the outset of the lesson. After students got their books ready for dismissal, she asked them hurriedly, “What are words that mean addition?” Students quickly repeated the expected words. She asked, “What are words that mean subtraction?” Again students repeated the key words. She said “Excellent. You may go.” (Her lesson closure emphasized rote recall.)

I wrote the following reflection after my observations that day:

After class when Ms. Drake had a break from some departmental responsibilities, I said I had a question about the class. She said, “Isn’t it unbelievable how poor their vocabulary is?” As usual when she asks me this kind of question about the class, I sort of shrugged, looked surprised, and said that it was hard for me to tell from where I sit. In truth, I’m not sure what about the class’ actions tell her that. They don’t ask questions. (Maybe she sees their facial expressions.) She went on for several minutes about how sad it is that their skills are so weak. She asked if I noticed how she didn’t go on to evaluating expressions? (I had.) She felt they needed lots of time working on the vocabulary. But for the most part, it was all quite isolated. She did use some contextualized expressions such as “The distance, D , decreased by 8 miles,” but there was no connection made to any realistic equations. She said she had “learned to assume they know nothing.”

I asked her about the tiles for $x - 3$. I said I knew that the kids did not yet learn integers, but isn’t $x-3$ in tiles the same no matter where you put the tiles? She said, yes, but she wanted them to learn that $x - 3$ and $3- x$ are different. She needed to reinforce that. She said she was using the tiles “strictly as a visual... So the kids have something to visualize.” She said that the variable as a letter is totally abstract, so now they have the rectangle to think of. (I don’t understand her reasoning. She never used the tiles to reinforce the idea that “the letter”

stands for an unknown quantity, or a changing quantity.) She said when they do integers, that, yes, then with the tiles, 1 rectangle and 3 reds will be the same as 3 reds and then a rectangle. (I think this could be very confusing for the students because she said the red meant subtraction, rather than a negative unit.) (Ms. Drake field notes/post-observation reflections, February 22, 2000.)

In the fourth lesson, the new concept was evaluating expressions. She used very simple expressions, such as “ $x + 4$ ” and had students use the algebra tiles to create the expression and then substitute the rectangle (variable) with the given quantity. The problems they used were not especially context-bound. It is not clear why she thought tiles were particularly useful for either this lesson or the previous. She eventually had them move away from using the tiles and stressed using the “funnel” method, whereby they were to “work vertically down the funnel. Start at the top. Everything flows to the bottom [as they evaluate an expression.]” Most of the class time was spent practicing problems out of the book following the ten-minute lesson. One illustrative exchange took place during this lesson that demonstrated the kind of interaction that often took place between her and the students. Students had trouble evaluating $20 + 4.5$. One student said 60.5 and began to say “I was thinking...” She cut him off and said rather loudly, “You thought? Don’t think! You were thinking wrong!” She reminded the class about lining up the decimals without mentioning place value or discussing number sense.

Algebra II (GT)

The Algebra II class that I observed Ms. Drake teaching was the only Algebra II eighth-grade class at the school. The students in the class were referred to as the “GT” students -- “gifted and talented.” I will herein refer to this class as the GT class. Ms.

Drake had also taught these students in seventh grade, and she said that that class had started with over 30 kids. Most weren't "true GT," so that they were whittled down to these 14 students. I observed 8 classes. Table 20 shows the three-week span over which I observed.

The structure of the GT class periods was similar to that of Fundamentals. That is, every class started with a drill that lasted from four to nine minutes. Answers were given by students or put on the board; there was very rarely any oral discussion about the drill. Drill exercises were in multiple choice format (modeled after the High School Assessment problems that were to be piloted in the district this year) and seemed to be review problems on material they covered that year. On three of the days, homework answers were called out in less than four minutes with no discussion. On another three days, either Ms. Drake or the students put homework problems on the board because there were many problems students had been unable to do. On a few occasions, there were administrative tasks to be done, such as filling out their "table of contents" sheets or reminders about filling out their math logs. The remainder of the periods had either a whole-class lesson or activity, or pairs or individuals working at their seats. There was little if any closure brought to the end of classes.

During the eight days of observations, and during the "lesson" portion of the periods, the class completed a chapter on functions, reviewed some statistics topics, and began the chapter on exponential functions. I was able to see the introduction and development of recursive functions, the final topic in the chapter on functions. This development took less than two class periods. Ms. Drake's presentation of recursive

Table 20
Time Table for Observations of Ms. Drake's GT Class

| Date | Observation / No observation |
|------------------------|---|
| Thursday, February 10 | Observation |
| Friday, February 11 | No observation – class was “canceled” due to a field trip that left only three students in her class. |
| Monday, February 14 | Observation |
| Tuesday, February 15 | Observation |
| Wednesday, February 16 | Observation |
| Thursday, February 17 | Observation |
| Friday, February 18 | No observation – Snow day |
| Monday, February 21 | No Observation – President’s Day |
| Tuesday, February 22 | No observation – Test given on functions |
| Wednesday, February 23 | Observation |
| Thursday, February 24 | Observation |
| Friday, February 25 | Observation |

functions was typical of her instructional practice for developing concepts. Below I will describe in detail those two days of lessons, as well as the comments I made in field notes written immediately after the classes.

Ms. Drake began the first day’s lesson on recursive functions by giving the definition of recursive functions as “a function that repeats itself unless you set limits.”

Note this definition is incorrect. The text used for their course states:

A recursive function is a function whose domain is the set of nonnegative integers (or sometimes the set of positive integers). To indicate that the domain is this set and not the set of real numbers, n is usually used as the variable. A recursive function can be defined by stating the value of the function at 0 and by giving an equation for the value of the function at n in terms of the value of the function at $n - 1$. (Larson, Kanold, & Stiff, 1993, p. 321)

Ms. Drake then gave examples of factorial usage using combinatorics problems. “In the lottery, you want three numbers. Your numbers are 3, 4, and 5. How many ways can those numbers appear?” (She did not explain the relationship this example might have with recursion.) Students sat in silence watching her. After presenting a second combinatorics example, she said, “That’s one example of a recursive function.” It was unclear as to whether Ms. Drake was referring to the second example or whether she was characterizing all combinatorics problems as recursive functions.

She then said that they would look at some geometry examples. However, as if she remembered she wanted to do something else first, she wrote a recursive problem on the overhead: “Find the first five values of the recursive function $f(N) = f(N-1) + 2$; $f(0) = 2$. They went on to find $f(1)$ through $f(4)$. (She never related this exercise to her definition of recursion. The students seemed to learn to do what she did, as she substituted and evaluated the expressions.) After they evaluated the four values, she asked them, “What’s the pattern?” Students said they were getting the even numbers. She did another problem that was similar, in which after evaluating $f(0)$ through $f(3)$ for a different function, students were asked to recognize the pattern of output values. She still had not pointed out the recursive nature of the defined functions.

The next 18 minutes was spent beginning and finishing a potentially rich geometry activity connecting many concepts and skills they had learned, including a tie to recursive functions. The problem essentially asked students to determine the number of diagonals there would be in any regular n -gon. The title on the handout said, “Geometry on the TI-82” indicating that calculators were to be used as a tool. The handout showed drawings of five different polygons asking, “How many diagonals does

each of the following polygons contain? Complete the table below.” The table had three columns with the headings “Name of Polygon,” “Number of Sides,” and “Number of Diagonals.” The seven rows were blank except for one polygon name – dodecagon, with one more row beneath it. Although this activity might suggest that the teacher be explicit about what the intended goal for the students was, which was to extend the table to include an n -gon, this activity was also enhanced by its open-ended nature. That is, students could work to find a pattern that defines a function explicitly, or they might see the recursive pattern. They might use the graphing calculator or they might not. If they did, they could use the table features or the statistical features. Ms. Drake did not elicit any ideas such as these, and as the observer, since the topic of the day was “recursive functions,” I assumed this activity would tie into that notion. The description and commentary below is taken directly from my field notes written immediately after class. It illustrates how Ms. Drake presented the activity in a very controlled manner, having students do just what she directed them to do. There was virtually no explanation as to why various steps were taken.

Ms. Drake began by defining a diagonal as “a line segment that connects 2 vertices of a polygon, but can’t be a side.” Then she completed the first three polygons in the table. She had students do the next two. Then she said that a 12-gon was too hard to draw so she immediately went to the stat mode of the calculator to plot the relationship between sides (x) and diagonals (y). She did not elicit this strategy, nor did she state the relationship she was apparently looking for. She did not explain her idea of using the 5 points to find a graph of a function to then find $f(12)$. She gave directions all the way through using the stat keys. She did not act like the students were seeing it for the first time, but she didn’t ask them what to do, either.

Once the points were plotted she said, “That’s not going to help us figure if it’s linear or quadratic. Let’s go to equations.” She immediately went to the quadratic key under the statistics menu and obtained the parameters, a , b , and c for the plot. Next she wrote the equation into the “ $Y=$ ” window and graphed the function. She said, “See, it’s a perfect line of best fit.” (She did all of the calculator manipulations in a very robotic way. She barely ever paused to see if

anyone was with her and seemed to be talking herself through the steps. She never made any attempt to connect the big ideas in this lesson.)

After finding the line of best fit, Ms. Drake asked, “How can we find out how many diagonals are in the dodecagon? We can substitute 12, but it’s easier to go to the table.” She went to the table mode, found that 12 sides implied 54 diagonals, and filled it in as such in the table. She also filled in the last line in the table with “N” sides, and the number of diagonals as $\frac{1}{2}N^2 - \frac{3}{2}N$ without any discussion of what she had found. Then Ms. Drake asked, “Now what’s the pattern?” (Presumably she meant from one y to the next as x increases by 1, but she didn’t talk about the difference between the explicit function already identified and the pattern she might be asking for.) One student saw the pattern and said, “As the number of sides increase the number of diagonals increase in numerical order.” This response was not questioned or clarified further as to what the student meant or how to write what she said symbolically (recursively). (There was no discussion of the function as an explicit definition for the recursive pattern. Usual recursion notation was never shown.) Ms. Drake told students to predict the number of diagonals for an octagon and the lesson ended abruptly. The geometry activity took a total of 18 minutes.

This activity had potential for rich mathematical discussion about patterns, functions and recursion, but Ms. Drake did not capitalize on those opportunities. There had been no discussion of the quadratic function they “found” even though the class had just finished studying transformations of quadratics the previous day.

Ms. Drake took control of this lesson and she left little room for questions. Although she occasionally asked if anyone had any questions, students rarely asked any other than “What was that number?” (Ms. Drake field notes/post-observation reflections, February 14, 2000.)

Neither at the end of that class period, nor the next one was there any mention of the relationship of that geometry problem to recursion. At the very end of class one student had named the recursive pattern she saw, but this was not written symbolically, nor was it related to the functional (explicit) definition that the calculator gave them. There was no looking back to see if other students saw the recursive pattern. At break the following day, before the GT class, Ms. Drake told me she was happy that “the students, well at least Sarah, saw the pattern.” This was after I’d asked her what she’d hoped the students would learn from that lesson. She said she really just wanted them to

be able to recognize and name a pattern. She said, “That stuff is on the MSPAP and these kids are the ones who help our scores.”

The day following the first geometry problem, Ms. Drake presented another geometry activity that could have taken two periods on its own. She spent only 20 minutes on it, whereby it served to close the topic of recursive functions and then spent the last 12 minutes of class getting students started on a performance task on measures of central tendency. Below is my description and commentary paraphrased from my field notes from that day.

Ms. Drake executed the second geometry activity in nearly the same manner as she had done the first geometry problem the previous day, although with this second geometry example, she never mentioned the word “recursion.” This problem asked students to find out how many chords could be drawn on a circle, given “ n ” points on the circle. After going over the definition of a chord, students filled out the chart for the number of chords that can be drawn, given two through six points on a circle. Then the students watched the overhead as Ms. Drake led them through finding the graph and function for this problem using the graphing calculator. The class never addressed the recursive pattern evident in this problem.

On the overhead, Ms. Drake put the data in the “lists” window of the calculator. Students followed on theirs. She asked them, “What kind of graph is this?” Students only had a table in front of them -- no graph -- yet she was expecting them to tell her what kind of graph this would be. (There was no mention of differences, which might be a way for students to tell.) One student replied “quadratic.” (Ms. Drake told me later that the student had probably graphed it ahead of time; it could also be that the student

said quadratic only because that had been the case on the previous day.) Ms. Drake then set the graph window. In the “window” screen she went through each setting out loud, but seeming to talk to herself. When the graph was too “flat” she said, “Let’s change the window.” She repeatedly put in y-max values that were too high so that her points were not spread out vertically enough. She never asked the students for help, nor did she mention what she liked or disliked about her windows.

Then Ms. Drake asked, “It looks like its going to be what kind of line?...Angie told us quadratic.” She did not elicit any other ideas, nor did she ask why they might think that. (Because her graph was so flat, it was not really possible to tell that it would be quadratic.) She used the calculator to obtain the parameters and immediately went to the “y=” window to graph the quadratic. The line indeed seemed to go through the points. From the graph the class was able to obtain the answer to the number of chords given 20 points on a circle. In the last line in the table, Ms. Drake wrote “n” in the input column and “ $\frac{1}{2} n^2 - \frac{1}{2} n$ ” in the output column, and she said, “That’s a nice quadratic.” The lesson ended abruptly at this point, only 20 minutes after the introduction.

Ms. Drake then announced that they were “switching gears” to do a performance task on mean, median, mode, and box and whiskers plots that would help them prepare for the MSPAP. She handed out the task, but then had them read in their books for several minutes to refresh their memories about the definitions involved. Then she presented the first question aloud using the overhead calculator. She briefly stumbled when the plot she chose gave her an outlier outside the “whisker.” After some time, she realized it was an outlier, and that she could have chosen a different option so that the whisker would reach as far as the extreme range. During the moments when Ms. Drake

was flustered and not sure what that “box” was hanging to the right on the screen, she never elicited ideas, and never mentioned the idea of looking at the data to check their graph. When she figured out what had happened, she discussed outliers briefly and then switched graphs. She did not explain that even though they chose a different display, the outlier was still there. The lesson ended abruptly, as class time was over. Ms. Drake told the students to draw the three other box and whisker plots for homework. They never actually worked individually or in groups on this problem during the period.

Ms. Drake had told me before the class that she was planning to do some statistics with them because they were at the end of a chapter and there was always a statistics review at the end of chapters. She added that it was getting near MSPAP time and she wanted to make sure the GT students would be prepared since “they’re our only hope.”

Ms. Drake’s Instructional Practices

Tasks

In the fundamentals course the tasks in which students were engaged were overwhelmingly working practice problems from the book, after observing an example done by Ms. Drake. For the classes in which I observed, no activities were used that might have provided opportunities to develop conceptual knowledge or to connect or extend knowledge. The lessons did not appear to begin with any particular motivational strategy in mind, and I observed certain sequencing problems. For example, algebra tiles were used in a demonstration before explaining their representational characteristics. Although Ms. Drake attempted to use algebra tiles to illustrate another

mode of representation of symbols, their use appeared more to hinder, rather than facilitate, clarity concerning the topic at hand. Additionally, Ms. Drake typically gave students a review sheet on precisely the same material as would appear on a quiz just prior to giving the quiz. Both the review sheets and the quizzes had nothing more than problems similar to the ones in the book that students had done for classwork and homework and served to reinforce the notion that mathematics is a matter of practicing rules, definitions and procedures.

In the GT class, I observed Ms. Drake using several activities that had potential to foster conceptual understanding in a variety of topics. However, for the most part, the tasks were strictly guided by the teacher, with little or no student input, and permitted no room for student exploration of concepts. When concepts and activities were introduced and developed, the connection between concepts and activities remained obscure; closure was non-existent. Connections to related concepts were not elicited or discussed at all. For example, at the conclusion of the activity on chords, the class had obtained an explicit function by having the calculator fit a quadratic. However, the function as they wrote it did not make the recursive pattern in the data evident at all. This pattern might have become obvious to the students had they looked at their data in a variety of other ways.

Calculator usage was common, but was carried out in a very procedural manner, with no exploration of some of the difficulties in working with such technology. For example, several times there were opportunities to discuss the limitations of the graphing window due to the finite nature of the pixels. Unfortunately, when something unexpected occurred on the calculator, Ms. Drake tried to resolve the issue before any

discussion could occur. It is noteworthy, however, that Ms. Drake pulled these activities from sources other than the main textbook. In one case, she mentioned that she would use a calculator exercise that “she was told to use.” I do not know whether or not she chose to do the others because she, herself, found them valuable, or she felt pressure to use them from outside (district or regional) sources.

Learning environment

Ms. Drake’s strongest quality was her rapport with students, as she commanded respect and expected respect for others in the classroom. Her demeanor was almost always very calm, but directive. Students almost never acted out in inappropriate ways. I once saw her ask one girl to switch seats with another because one of them had been talking with her neighbor. The two girls switched seats immediately, with no commotion or protest. She had very well established and seemingly effective classroom management skills. For instance, she had routines set up so that there were never any petty discussions about missing pencils, books, etc. She had several electric sharpeners in the rear of the room, and students knew to come in and sharpen their pencils before they sat down. If students did not have a pencil, she had pencils “for rent” on her desk, which students knew to borrow as they walked in, leaving something of theirs as “collateral” in exchange. There was no need for discussion of these matters. She used the “table of contents” sheet in both classes to help students keep track of everything they had done and to know what they were responsible for turning in. She frequently had them take out the sheet to record items as she dictated them. She then collected notebooks on occasion and checked that all items were included.

The downside to the learning environment that was established was the controlled manner in which she ran the class. Time was segmented carefully to keep things moving. All desks faced front so that interaction with other students was minimized. She did not encourage students to explain solutions, either to her or to each other. In the GT class she did have students work in pairs on several occasions but did not generally state the goals of the pairs other than to come up with answers together or to “compare answers.” On one occasion, she did tell pairs to “discuss and argue,” but when the whole group came together, the discussion was limited to answers only. The learning environment did not foster a drive toward conceptual understanding or the idea that mathematics involves sense-making.

Discourse

The discourse in both classes was extremely limited. The questions that Ms. Drake posed to the classes nearly always required only the answer to a mathematical calculation. She did not press correct answers for justification or explanation, nor did she probe wrong answers. She made no attempt to use questions that would probe understanding, or that would engage or challenge students’ thinking. Thus, there were nearly no exchanges beyond question-answer, whatsoever. In the few cases when she asked a conceptual question, she often answered it herself, having given almost no wait-time for a response at all.

Likewise, Ms. Drake did not expect students to question one another. In whole-class learning, nearly all exchanges were teacher-student. As mentioned above, only in the GT class did she have pairs work together, and due to where I was seated, it was

difficult to know what kind of exchanges took place. However, because she did not emphasize conceptual understanding, it is likely that students only questioned each other on procedures or answers.

Other themes

Other themes that were evident in the data collected on Ms. Drake's instruction were her level of content or pedagogical knowledge, her attitudes toward students, and her attitude toward the state performance assessment. Each of these will be commented on below.

Content knowledge

The lessons in the GT class described above serve to illustrate that Ms. Drake had a rather superficial understanding of at least some of the content she taught. Her "lecture" on recursion contained inaccurate definitions and poor examples. Limited mathematics content knowledge makes pedagogical decisions difficult. This could be one reason why she was unable to connect the activities she chose to do with the crux of the topic. The introduction to algebra lessons described above also illustrated her inability to motivate and conceptually develop fundamental ideas in mathematics. She appeared to have a very narrow, skill-oriented, view of mathematics. When she talked about assessing students, she mentioned only skills. She never used the words "concept" or "understanding." The result of her limited mathematics content and pedagogical knowledge translated into the traditional style of instruction and controlled learning environment she created. Artzt and Armour-Thomas (1999) validate this notion:

One might say that a teacher-directed style of teaching can serve as a mask for teachers who do not possess full knowledge of the content, students and/or pedagogy. That is, without the demands arising from student input, teachers are free to impose the material on the students even when they themselves do not fully understand it or have inappropriately sequenced the material (p.229).

This view of Ms. Drake was further bolstered by many comments she made to me over several occasions, either as unsolicited comments, or prompted only by my asking her a question about a lesson I had just observed. These comments spoke to her insecurities about content or “new” pedagogical practices she felt pressure to emulate. For instance, she once told me that she took a Pre-Calculus course given by the district for one week during the past summer. She and some teachers took it, thinking about moving up to a high school. She said that “the trig blew [her] away – it’d been too long.” Even more to the point, a notice she received that required her attendance at a meeting with the vice principal and principal to discuss reading and writing to assess students caused her to be very upset. The notice prompted her to make the following comments to me during the break:

I’ve gotta make sure they know their basic skills before I can take them where they need to be...Let’s face it – how many of those kids are going to four-year schools?...You’ve gotta go over how to do the example from the book. They can’t follow it on their own...I just don’t know if I’m doing what I’m supposed to be doing...I just want to be able to come in and teach the way I wanna teach...There’s so many skills they need to learn. I’ll fit in reading and writing when I think it’s appropriate...I know from what they put on the board and from walking around looking at their work what they know and don’t know.

After a short conversation on another topic, she came back to her previous thoughts, and said that back when she was in college, “the teacher just walked in, went over homework, did some new examples. That’s it. [She] got it. There was no problem.”

The next day during break, she made similar comments, lamenting about how the rules had changed in middle schools, so that they no longer had such luxuries as end-of-period bells or tape on the floor to line up desks. She said, “These kids get no discipline at home. They *want* discipline here. They don’t want chaos in the math class.”

Attitude toward students

In some ways Ms. Drake clearly appeared to care about her students. She spoke about many of their home problems with sincere worry and concern in her voice. She seemed very well-liked by students, as evidenced by their informal conversations with her just before class officially began or after dismissal. She was both the boys and girls basketball coach, and seemed to know many of her students through that activity or from having them in previous years.

At the same time, she made many comments to me indicating that she felt like she was in an inferior school with inferior students. At least twice during my observations, she told me it was a shame I didn’t get to observe in [a different region of the district.] She didn’t use specific words, but was hinting at her students just not being as good a crop, and “who knows what jobs some of these kids are going to be able to get.” She seemed especially annoyed at having two students in her Fundamentals class who had come from another district where they had been “inclusion kids.” She would ask me, “Did you see that one? He doesn’t talk; he just mumbles.” Several times after class she would ask me if I saw what she meant about those students. I would say that I honestly couldn’t see anything from the back of the room. Yet, I was thinking that I

wasn't sure where she got her impression because so little was asked of students, and she did not try to interact with those two students, in particular, at all. More than once she referred to students in that class as "dodos" or "bozos;" she once said that on the following day students would be using the calculator they would need for the state performance assessment exam because it was different than what they had been using and "these dodos will never figure it out."

Attitude toward PA

The first time Ms. Drake mentioned the statewide assessment to me, she said, "How 'bout that [PA] test? I've been giving that test from the beginning and haven't seen the sense in it yet. It's made for GT kids." On other occasions she commented again about how ridiculous the PA was, and how she needed to try to prepare the GT kids for it, since they were the school's only hope. She never stated exactly why she felt that way. She only insinuated that it was just too difficult for non-GT kids.

Comparison of Ms. Drake's Fundamentals and GT Classes

During the course of my observations, Ms. Drake used more activities that had the potential for integrating various concepts and topics and that were more open-ended in nature with her GT class than with the Fundamentals class. In the Fundamentals classes that I observed, topics were presented in a more isolated manner, with no extended or open-ended tasks provided. However, as discussed above, the tasks in the GT class were implemented in such a way as to essentially eliminate the opportunity for students to make over-arching connections among mathematical topics. That is, Ms.

Drake controlled the direction of the activities in such a way that did not leave room for students to think through the activities prior to her demonstrations. Because of the way in which students completed the activities, it is very difficult to know what conceptual knowledge the students gained. However, whereas these activities used up the bulk of class time for the GT students, the Fundamentals students were quizzed much more often, and were asked to work on problems out of their textbook individually during the lesson time. The problems they worked on were unquestionably more routine and compartmentalized than the kinds of tasks to which the GT students were exposed during my observations.

There were other noticeable differences between the GT class and the Fundamentals class. The GT class used the graphing calculator often, whereas only during one of my observation periods did the Fundamentals class use any calculators at all. However, this can be attributed to the difference in content that the students were studying at the time. Moreover, although the graphing calculators were used as tools to accomplish various tasks (finding line of best fit, drawing statistical graphs, graphs of functions, looking at tables) as previously discussed, very little discussion took place. It is unclear as to whether the students were able to tie together the significance of the calculator's findings with the mathematics they were doing. The Fundamentals class used calculators to work on order of operations on a day I did not attend class.

The GT class was considerably smaller in size (14 compared to 23), which could provide that class the advantage of allowing for more individuals to have a larger part in the discourse of the classroom. However, discourse was minimal in both classes. Ms. Drake also asked the GT class to work in pairs on several occasions, something I did

not observe in the Fundamentals class. The pairs often were put together for the purpose of checking their homework, but it is difficult to tell what kind of discourse took place among the pairs.

Ms. Drake was fairly explicit regarding her strong doubts about the abilities of the students in her Fundamentals class, whereas she had personally made sure that this GT class was made up of “true GT” students. This difference in attitude was made clear on numerous occasions through explicit verbalizations.

PA Alignment with Instruction and Opportunity to Learn

Ms. Drake made it clear to me that she was aware and concerned about the content of the PA, as evidenced by her including a review of statistical concepts and techniques for the GT students during my period of observations. The activity she did with them had many of the elements of a “worthwhile mathematical task,” as described by the NCTM Professional Teaching Standards (NCTM, 1991). That is, the activity they were given had the potential to “call for problem formulation, problem solving and reasoning; promote communication about mathematics; stimulate students to make connections and develop a coherent framework for mathematical ideas; engage students’ intellect; [and] develop students’ mathematical understandings and skills” (NCTM, 1991, p. 25). Although students did some of the work for this activity in pairs, the subsequent whole-class discussion failed to generate interaction focusing on concepts and differing opinions or perspectives. Ms. Drake seemed to redirect answers to match the answers that she had in mind. In nearly all cases, she did not probe answers that were different. There were no other activities that were directed explicitly at

preparing students for the PA. As described above, Ms. Drake did use several other activities that would be deemed “worthwhile mathematical tasks” and inherently ought to be aligned with expectations of the PA. However, because of the way in which they were completed, it is unclear whether the students would be able to independently formulate a plan of action for a problem-solving situation.

I did not observe Ms. Drake doing any “worthwhile mathematical tasks” during the Fundamentals class, and in fact she did express to me that she felt the GT students were “their only hope” on the state PA. The lessons in the Fundamentals class that I observed did not align with the state’s expressed instructional expectations for the PA.

Ms. Henderson

Ms. Henderson was teaching at a middle school whose total enrollment was over 900. The minority enrollment was not quite five percent, the second lowest minority representation in the district’s middle schools. Of students receiving free or reduced lunches, this middle school had the lowest percentage in the district, at only five percent. I observed Ms. Henderson for three weeks beginning March 13, 2000. I observed two classes back-to-back. The first was a Pre-Algebra course, and the second was one of two Algebra II courses in the eighth-grade at that school. The front of the room had a large chalkboard and a pull-down screen and there were two long chalkboards on both sides of the room. The arrangement of the desks is pictured in Figure 5. The circles depict the location of the 27 Pre-Algebra students. The room was arranged in that manner for all classes observed in both courses, except for one day in Pre-algebra when the desks were re-arranged in clumps of five, presumably to provide for a more conducive atmosphere for that day’s group activity.

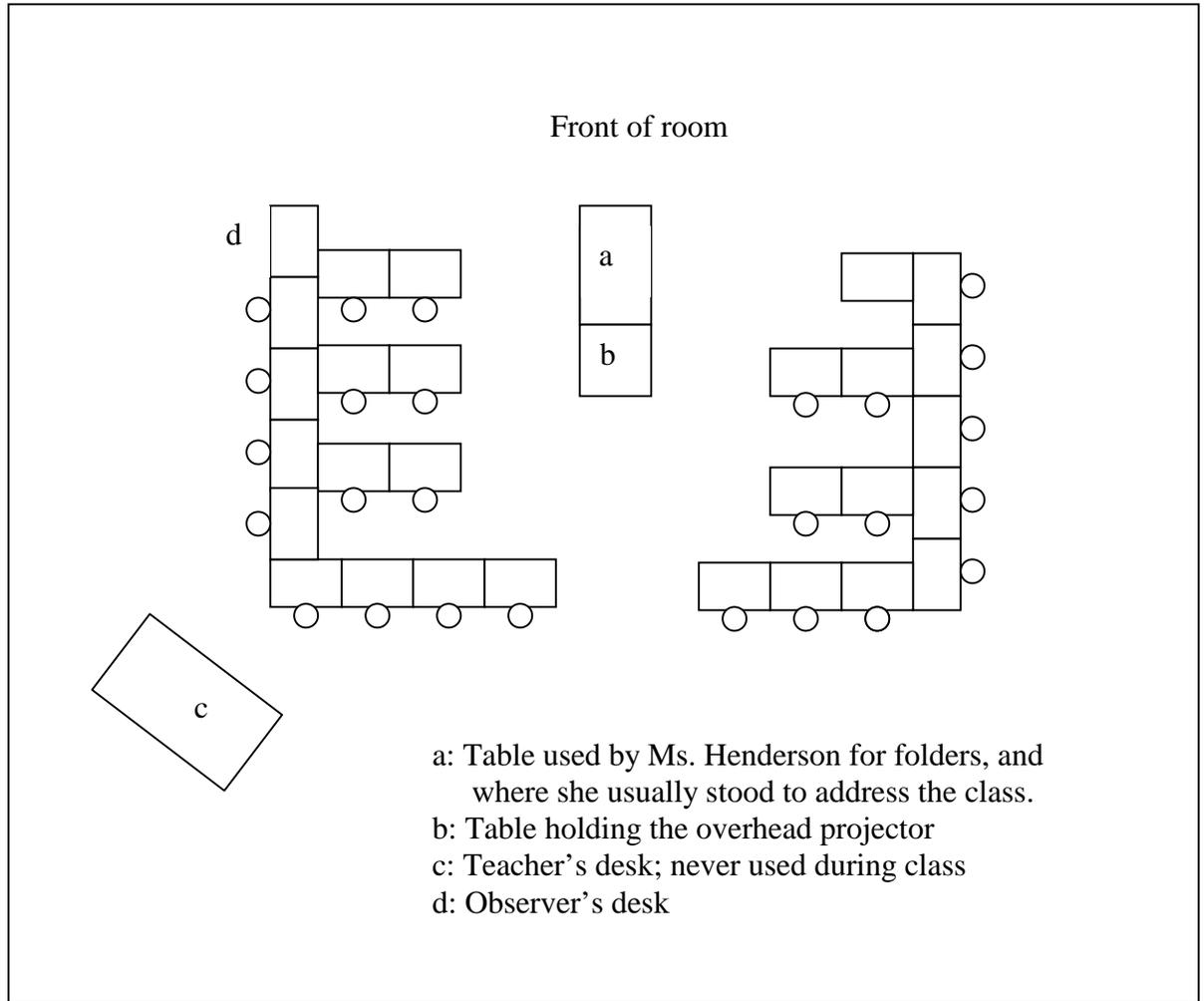


Figure 5. Ms. Henderson's classroom.

Pre-algebra

The Pre-Algebra class had 27 students. They sat in a location pre-determined by Ms. Henderson twice per quarter so that they were next to a partner and also in a group of four. Ms. Henderson explained to me once after an observation that she put them in groups according to quarter grades, mixing highest with lowest, taking into account personalities. Thus, without their necessarily knowing it, students were grouped rather

heterogeneously. I observed the class 12 times. Table 21 shows the timeframe over which I observed, and whether an observation took place or the reason one did not.

During the timeframe in which I observed, the class reviewed and took a test on solving linear equations and inequalities and then started a geometry unit. Much of the geometry unit was review from the year before, but new theorems and definitions were also introduced. I observed the class going over basic geometry terms, learning about particular angle relationships, measuring angles with protractors, and doing a number of constructions using patty paper and traditional compass and straightedge constructions.

Ms. Henderson's lessons generally followed a particular format. As students walked in, there was an "opener" problem on the overhead or on their daily handout. The opener was generally done in two distinct segments. The first several minutes allowed time for students to complete the opener either individually or with their partners or groups. During this time, homework was also out on their desks for Ms. Henderson to check, and sometimes homework problems were put on the board during this beginning-of-class time. Next, the class would discuss the opener in detail as a group. The entire opener took between 6 and 23 minutes, with an average time of about 12 minutes over the 12 classes that I observed. After the opener discussion, homework was discussed as a whole class. The teacher generally posed questions to check for understanding and to determine whether or not students had questions on the homework. This took anywhere from 5 to 15 minutes. She occasionally collected the homework, particularly on days when she could not spend a lot of time discussing it. The next lesson segment was the main part of the lesson, in which new material was developed or presented. This portion usually took up 30 minutes of class time

Table 21
Time Table for Observations of Ms. Henderson's Pre-Algebra Class

| Date | Observation / No observation |
|---------------------|---|
| Monday, March 13 | Observation |
| Tuesday, March 14 | Observation |
| Wednesday, March 15 | Observation |
| Thursday, March 16 | No observation – teacher's suggestion. Class took full-period test. |
| Friday, March 17 | Observation |
| Monday, March 20 | Observation |
| Tuesday, March 21 | Observation |
| Wednesday, March 22 | Observation |
| Thursday, March 23 | Observation |
| Friday, March 24 | No observation - class played a game that would be highlighted at upcoming Parent's Night |
| Monday, March 27 | Observation |
| Tuesday, March 28 | Observation |
| Wednesday, March 29 | Observation of MSPAP simulation – regular classes canceled |
| Thursday, March 30 | Observation |
| Friday, March 31 | Observation |

and generally included a short closure segment. Since Ms. Henderson had clearly distinct segments to her lessons, I will describe Ms. Henderson's instructional practices by discussing each of those main segments, using specific examples from lessons to support the interpretations. I will then describe one lesson in more detail to illustrate how she connected one lesson to other lessons, as well as how she tied components of a lesson together.

Ms. Henderson almost always had a daily handout that was titled and contained the “Opener,” the “Goal” for that day, a “Class Discussion” section, and/or “Class Demonstration,” and/or “Classwork.” The opener was either a problem that reviewed the previous day’s material and simultaneously served to segue into the new material to be presented, or the opener acted as a motivator and served to bring out conceptions of the new material that students brought to class. In the former case, the opener generally acted as an extension problem, rather than something that mimicked what they had done the previous day. The openers were done either individually, or in discussion among partners or groups, a decision apparently left to student discretion on any given day. In examining answers to the opener problem, a student volunteer was selected to come to the front of the room in the capacity of a “recorder,” to summarize in writing on the overhead the ensuing class discussion.

During this whole-class discussion, Ms. Henderson used questioning constantly, in order to make students articulate their answers and reasoning. She consistently asked, “Why,” and “How do we know...” to clarify as much conceptual knowledge as procedural knowledge. Even when the opener was review, she still went over its key concepts in some depth, rather than just stating a rule. For instance, in an opener that was review of past material and occurring a day before a test, students were given a series of linear inequalities and had to determine whether the inequality symbol would be reversed in the course of finding a solution. Although she had many students explain why the sign would or would not be reversed (referring to the rule in their explanation), she also went on to give a conceptual argument for *why* the rule is what it is.

Ordinarily, homework was discussed in depth, although a few times

Ms. Henderson told the class that time was short so that she would collect homework, rather than spend too much time discussing it. When homework was discussed, she used it as an opportunity once again to highlight important mathematics by asking whoever was answering a particular question a series of follow-up questions to make sure that the original answer was justified. Ms. Henderson rarely let an answer be stated without an explanation or follow-up question. For example, when a student named an angle as an answer to a homework question, she followed with, “Why can you not just name an angle by its vertex?” She often addressed these questions to a particular student in the class, not necessarily the person who answered the question, nor other student volunteers.

The main portion of class time generally took about 30 minutes, and usually included a recognizably distinct segment that served to analyze the problem at hand, elicit ideas, and motivate. For example, on March 14, which was “3/14”, Ms. Henderson and the other Pre-Algebra teachers planned to do a lesson on the relationship between circumference and diameter of a circle in honor of “Pi day.” (They did not tell the students this until after the lesson.) The main part of the lesson involved students measuring various objects, filling in a chart, recognizing the constant relationship between C and D, and also that the constant is slightly more than 3. This lesson began with a whole class discussion of the terminology, and then a demonstration. During the demonstration, one student volunteer came up to use a string to see about how many times the diameter would fit around the circumference of a soda can. Ms. Henderson asked questions to help students express the relationship they were seeing as an algebraic equation. When one student offered a suggestion, she asked the class, “What

do you think about that,” without letting on whether any of the ideas were correct or not. Before letting the students try various measurements, she first made the goal clear – to get a better approximate than 3 – and had them read through the “Procedures” section of the handout circling the verbs to help them comprehend the directions of the task.

Ms. Henderson’s Pre-Algebra lessons generally had a main section in which tasks were completed. Of the 12 classes of Pre-Algebra that I observed, students had actual hands-on activities during 7 of the days. Some of these were classic geometry constructions; others were activities aimed at helping them learn a skill (measuring angles with protractors) or tasks that asked them to follow directions to review and connect various concepts they had recently encountered (cutting a rectangle into a trapezoid and triangle, then measuring the angles, and identifying angle relationships.) Such tasks had clearly stated purposes and were rich with potential to connect new knowledge to prior knowledge.

On days when no hands-on tasks were completed, the main lesson segment had either a teacher demonstration or a critical problem and was dominated by question and answer, whole-class discussions. For example, during a “lecture” to introduce the new concepts of supplementary, complementary, adjacent and vertical angles, Ms. Henderson asked a myriad of questions to encourage students to articulate new relationships:

“Do you know what the relationship between those lines is?”

“What did it ask you to do?”

“Why is it not obtuse?”

“How many points are on a line?”

“Can I have an estimate of this angle [gives sketch below]... Then what is this [points to supplementary angle]?”



“Tell me what it means for two angles to be supplementary.”

“What do you know about this angle from the picture?”

“If that is true, then what’s [this angle]?”

“What do these two angles share?”

“How did you know it was 115° ?”

By constantly asking questions, Ms. Henderson was essentially developing concepts by eliciting student ideas, rather than simply presenting information. This approach is posited as helping to foster a deeper understanding of concepts. Although there were times when Ms. Henderson might have included more wait-time after posing a question, she usually allowed for sufficient time, waiting to see at least several hands before calling on someone.

Pre-Algebra lessons often ended with some kind of closure segment in which Ms. Henderson focused the students on the main point of the discussion/lesson. For example, one day was spent continuing to work on solving equations and inequalities. The main segment of the lesson was initiated with a whole class discussion; then students worked on problems in pairs; then the class met as a whole again, discussing those same problems on the board. Four minutes before the period ended, she had students stop what they were doing, and individually write a response to the question, “How is solving inequalities different from solving equations?” After two minutes to

think and write, she had students share their responses aloud, pointing out the many different ideas that the students offered.

As an example of how Ms. Henderson's lessons built on one another and how the components of the lessons (including the opener) fit together, consider the following lesson. In the lessons preceding it, students had learned about straight, complementary, supplementary, adjacent, and vertical angles.

The opener on this particular day had six different diagrams with lines intersecting in various ways. One angle in each diagram was labeled with a measure, and the students needed to identify the measure of some other angle in the diagram without measuring. In this way, students were expected to use what they knew about angle relationships. She gave students six minutes to work on the six problems. Some of the problems had two unknown angles in the sketches. (I noticed that some students worked through this in their groups or with a partner; apparently, this was acceptable.) When the time came to discuss the problems, Ms. Henderson told the students to exchange papers with their partners and mark them. (This was not typical.) As the class discussed the opener, working through each angle, Ms. Henderson had the student-respondent review all of the concepts and definitions embedded in that particular problem. For example, she called on a student to give the answer to a problem. After the student gave the measure of the intended angle, she asked, "Why? Where did you get 154° ?" The student responded by stating that the angle under question was supplementary to the given angle of 26° , so that he subtracted from 180 to get the answer. In another explanation, a student read what his partner wrote. When she asked him how his partner got that answer, he said, "Actually he must have gotten this

wrong.” The teacher said, “Why do you think he’s wrong,” thereby asking the student to explain the angle relationships in that problem.

Next, Ms. Henderson conducted a whole-class discussion on intersecting lines as a segue to the parallel lines postulate, which was the topic for the day. She passed out a slip of paper and directed the students through a task to draw parallel lines cut by a transversal, soliciting their ideas for doing the steps along the way. She asked, “How could you use your ruler to draw a pair of parallel lines?” Several students said to use both sides of the ruler. She then said, “Now fold your paper in such a way that the fold will cross both lines, but not perpendicular.” She then asked more questions about how many angles they saw formed (eight) and asked them to use a protractor to measure all of the angles. (She wanted them to come to the conclusion that there were only two different degree measures among all eight angles and then use their prior knowledge to justify this outcome.) Ms. Henderson went around the room as they worked and said at one point, “Some of you are using those angle relationships to check the reasonableness of your measurements.” (This statement validated what some students did, gave other students a hint about the relationships, and reminded everyone that their prior knowledge gave them a way to make sense of their measurements.)

During the rest of the lesson, she used questions and answers to define corresponding, alternate interior, and alternate exterior angles and to point out the angle relationships within. For example, she began the discussion by asking, “How many different angle measures do you have?” A student responded by saying four. She probed him about this response, asking him about the measures of various pairs of angles, and asking him if they ought to be the same and why. After going through many pairs of

angles, each time having students explain the angle relationships, she remarked that there should only be two measures. When a student insisted that all of his angles measured 80° , she looked at what he did, and explained that he read his protractor incorrectly. She then lectured students about angles that are in the same position and defined corresponding angles, alternate interior, and alternate exterior angles by discussing the meaning of the words “alternate,” “interior,” and “exterior.” She then expected the class to decide which angles would be considered “alternate interior” and so forth. Students were then asked to try some practice problems on their handout that asked them to identify angle measures by using the relationships that they had just learned. Students spent only three minutes on this individually, after which time a recorder went to the overhead to write answers that other students gave. Ms. Henderson asked students to state the reasons for their answers each time. This sharing segment lasted two minutes and seemed to serve as closure, since all new terms were covered again within the context of the problems. After class, Ms. Henderson remarked to me that she had wanted to use patty paper (square slips of wax paper) but had been unable to locate it. She lamented that that would have allowed students to trace and see precisely that corresponding angles are equal.

Ms. Henderson’s Pre-Algebra class also spent three days on geometric constructions. Each day, students completed one classic construction, using two different methods. She had them try to figure out a construction on their own using patty paper and then led them through the compass and straightedge construction. Although these lessons also tied nicely into what students had learned, these lessons ended without as much clarity as many of her other lessons that I had observed. One

problem was that the compasses were rather difficult to use, and students were having trouble physically manipulating them to complete the constructions correctly. Students seemed frustrated and were clearly focused on getting the construction drawn, even during discussion time. Ms. Henderson's discussions of the constructions usually fell short of highlighting why the constructions worked. However, I asked her about this on the last day of my observations, after students constructed an angle bisector. She said that the purpose of doing these was essentially just to get the students used to the tools and the process of construction; they eventually would be taking an entire course in geometry where they would delve into the proofs behind constructions.

Algebra II (GT)

The Algebra II class that I observed consisted mainly of the eighth-grade students whose only GT class was in mathematics, as opposed to the other Algebra II class, which had students who were GT in all of their academic subjects. This particular class had only 16 students, which was small in comparison to the other GT class. The desks were arranged, and partners and groups were formed, in the same way as was described for the Pre-Algebra students. Table 22 shows the timeframe over which I observed 12 classes, and whether an observation took place or the reason one did not.

Just prior to my first observation of this GT class, the students had been studying quadratic functions and their graphs, as well as solving quadratic application problems graphically. The day I started observing, they began the second half of the unit on quadratics by studying the various methods of solving quadratic equations algebraically. These topics included solving quadratics using factoring, completing the

Table 22

Time Table for Observations of Ms. Henderson's Algebra II Class

| Date | Observation / No observation |
|---------------------|--|
| Monday, March 13 | Observation |
| Tuesday, March 14 | Observation |
| Wednesday, March 15 | Observation |
| Thursday, March 16 | Observation |
| Friday, March 17 | Observation |
| Monday, March 20 | Observation |
| Tuesday, March 21 | Observation |
| Wednesday, March 22 | Observation |
| Thursday, March 23 | Observation |
| Friday, March 24 | No observation – Students took a Partner quiz |
| Monday, March 27 | Observation |
| Tuesday, March 28 | Observation |
| Wednesday, March 29 | Observation of MSPAP simulation – regular classes canceled |
| Thursday, March 30 | Observation |
| Friday, March 31 | No Observation – Students took tests individually |

square, and the quadratic equation. (Solving quadratics by factoring was a review topic from Algebra I.) Students also solved quadratics by graphing, which was merely an extension of the graphing concepts they had already covered. During my observations, this GT class also addressed the concepts of imaginary and complex numbers, the discriminant, and worked on several application problems in which their goal was to solve a quadratic equation algebraically or determine if a solution was possible. I

observed them review for the test, which was given on the last day I visited that teacher. I did not stay for the test, although I received a copy of it.

The format of Ms. Henderson's GT class was very similar to that of the Pre-Algebra class. As in the Pre-Algebra class, the teacher usually gave students a daily handout that was titled and contained the "Opener," the "Goal" for that day, a "Class Discussion" section, and/or "Classwork." They began with an opener problem. While students worked on the opener, other students put homework problems on the board. The opener was discussed in detail, as were homework problems. The opener and discussion of it took between 13 and 28 minutes. Homework discussions lasted between 9 and 19 minutes. Because homework discussions tended to last longer in GT than in Pre-Algebra, the new concepts were generally developed in less time than those of Pre-Algebra, typically in about 20 minutes.

Opener problems in GT were related to recently covered concepts, and at the same time, usually acted as appropriate segues into the current days' topics. For example, on the day on which Ms. Henderson would derive the quadratic formula, the opener was a review problem requiring students to put a quadratic equation into standard form. The discussion of the opener began with Ms. Henderson asking the class many questions to draw out the important concepts or to assess their understanding. The answers to the opener problems were always made explicit after many questions and answers. These students seemed very concerned about understanding the algebra, as opposed to just learning rote procedures, as I commented in my field notes on March 22, "Kids ask good questions about the algebra." For example, after finding the two irrational roots of a quadratic equation, a student asked if it mattered if they kept the

sign on the denominator in the radical part of the root. The teacher returned this question to the students. Perhaps Ms. Henderson realized that the answer to the question depended on the form in which the roots were written.

Consider the following example of an opener that was used to check for understanding and helped tie together ideas. This opener was provided the day after students had learned about the quadratic formula, the last formal method of solving quadratics that they would learn. Students were asked to get into their groups, and each group was responsible for solving the same quadratic using a different specified method – factoring, completing the square, quadratic formula, or graphing. They were to put their solutions on poster-size paper. If they finished early, each student, on his or her own, was to solve that quadratic equation using all of the other methods for practice. In the end, the four student-solutions were hung on a side board and left there as a reference for students during the remainder of the unit.

Nearly every day, students in GT put homework problems on the board and were expected to go through their solution orally with the class. They were also expected to field questions about their solutions from both the teacher and the students. If a student asked Ms. Henderson a question, she would remind that student to address his or her question to the class.

The bulk of the new material each day was developed through whole-class question and answer discussions usually led by Ms. Henderson, but sometimes led by students. It may have been that this choice depended on how “new” the material was. (There were also days when students were able to use class time to work in groups on application problems, after beginning with a whole-class discussion.) The following

description of a GT lesson took place on the first day of my observations and illustrates many of the techniques and behaviors that Ms. Henderson employed regularly.

On this particular day, the opener problem was a review of factoring. The direction for the first expression said, “Write as a product of a monomial and a binomial,” and the direction for the next three expressions said, “Write as a product of two binomials.” (These directions, rather than just the typical direction, “Factor,” were instructionally sound, as they reinforced the meaning of factoring, which many students do not often consider.) Ms. Henderson told the students, “You have an opener to do. It’s a review from Algebra I. Let’s see what you remember about factoring.” After six minutes, she called on students to put the problems on the board. The students at the board explained to the class what they had done. After the first student explained, some students seemed confused. Ms. Henderson tried to clarify the student explanation by discussing the terminology of “terms” and “factors,” and she also mentioned “binomial” and “trinomial.” After that, she called on other students to state the terms in the remainder of the binomial or trinomial expressions, being sure they noted the positive or negative values. Ms. Henderson asked many questions to direct students through the factoring procedures, such as, “How did he know he’d need a ‘plus’ and a ‘minus’ [when he set up his binomial factors]?” She also asked them how they could check that a factorization was correct. All problems were explained very precisely by the respective student-solvers, and Ms. Henderson went over all explanations highlighting the strategies. The discussion of the opener ended when Ms. Henderson reminded the students to put the yellow reference sheet (on factoring) in the front of their notebooks.

Ms. Henderson then told the students to take out their homework from Friday, and said, “We’ve had a mixed review.” When she thought everyone was ready, she read the question, stating what the problem asked them to do, and then gave the answers. One student asked to see her solve a certain problem. She said she’d rather someone else do it. She reminded the student volunteer to talk her solution through as she wrote it on the board. (The problem was a quadratic application problem that did not require factoring.) Throughout this explanation, Ms. Henderson often asked, “How do we know...?” The homework discussion ended with Ms. Henderson saying that she would collect homework from any students who had questions about it.

Ms. Henderson then focused the class on the main part of the day’s lesson. (The goal for the day as written on their handout said, “We will factor quadratic polynomials and use the zero-product property in order to solve quadratic equations.”) Based on the daily handout, Ms. Henderson chose to use two application problems in order to introduce the use of the zero-product property, which was also review from Algebra I. Of the two applications problems, the first one was new to students, and the second problem was one that they had completed using the graphing calculator the previous week; now students would revisit this problem using an algebraic method.

Ms. Henderson began by giving students time to individually focus on the first problem. That problem concerned figuring out the width of a border for flowers around a rectangular garden, given a certain area for the garden. Students were asked to read the problem carefully, underlining important information. Then they were told to write how they would solve the problem, as Ms. Henderson asked them, “What would help you visualize the problem?” She gave them four minutes to analyze the problem

individually. Then she gave them two more minutes to share their thoughts and strategies with their partners. (Sitting amongst the students, I was able to hear that they used the time productively.)

After students were given time to think and share, Ms. Henderson asked for a student to draw a diagram on the board representing the problem situation, as she continued to ask questions of the class to assess their understanding of the problem such as, “What does it ask us to find?” As students called out to the student at the board to help complete the diagram using algebraic expressions, Ms. Henderson said, “You kids are approaching this problem very differently from me, so let’s see what we get.” (She apparently had a different algebraic solution in mind, but she was willing to let the students use their solution process, making sure at each step that they justified their decisions.) Eventually she asked, “So how can we write an equation for that area?” At this point, Ms. Henderson went to the board and wrote what the students directed her to write. Some students said they did not understand, so she went through and explained where each part of the equation came from, referring to the diagram. When students got to the point where they needed a strategy to actually find the solutions, she reminded them that in Algebra I they had learned the “zero product rule,” which required that the equation be set equal to zero. She led the students through the factoring part, soliciting their input at each step. She then asked leading questions such as, “Do you see a common factor,” and “Can we write this as a product?” After they obtain a factored expression set equal to zero, she moved to the side of the chalkboard, where she reminded them of the concept behind the zero-product rule. She began this by writing and saying, “If $A \cdot B = 0$, then $A = ?$ and $B = ?$ The only way two numbers multiply

together that gives a product of zero is for one of them to be zero.” (Her explanation sought to make sense of a procedure.) She then went back to their equation, which had factors set equal to zero and asked, “So what will make this [product] zero?” Students gave her the individual linear equations and then the solutions, which were 3 and -23 . She then said, “But this is a real-life application! Which solution is sensible?” She then reminded them to go back and answer the question asked in the problem. She reminded them that their homework did not have to do with this and mentioned things on the homework to look out for. (The class ran out of time after this, so that they completed the second application problem on the following day.)

Throughout the observations of Ms. Henderson, I was struck by how engaged the students were in discussions and tasks. They asked questions that could help them make connections or solidify concepts. For instance, a student once noticed a pattern when completing the square to find the vertex form for a quadratic. He asked, “Is the term in the binomial always half of the linear term in the trinomial?” The teacher replied, “Yes, but think about why that is. Can anyone explain?” Another time a student asked why the constant outside the factors has no effect on the solution to the quadratic equation – a concept easily overlooked by students and teachers. She explained the reason from an algebraic perspective, but this concept got brought up again later when the class got into a discussion over a particular quiz problem. That question had asked them to “solve by graphing: $-9 = 3x^2 + 12x$.” The teacher pointed out that many students solved this by finding the intercepts for and graphing the equivalent equation: $3 = x^2 + 4x$. She asked if they thought points should have been taken off of this solution on the quiz. Students engaged in a productive argument about what is the same and

different about the two equations after Ms. Henderson instructed them to graph the two relevant functions on their calculators. This discussion underscored a point that had been alluded to in several previous lessons, namely that the intercepts would be the same, although the graphs of the related functions were not.

In fact, very often in Ms. Henderson's class, concepts seemed to be cycled back, so that students had several opportunities to make sense of a concept or procedure. Teaching for understanding was also fostered by Ms. Henderson's many kinds of questions. She would ask, "What would we expect to get..." or "What would happen if I put (0,0) into the equation," trying to get students to anticipate certain notions. After deriving the quadratic formula by eliciting prior knowledge based on completing the square and other algebraic manipulations, she said, "That's it. It's the quadratic formula. Now what does it mean?" In this way, she asked them to "zoom out" and see the big picture. Her questions also focused students on relationships between concepts: "What is the relationship between the shaded region and the original inequality? What does the shaded region have to do with the graph?"; "How are they similar and how are they different?"; "Based on that solution, how would you expect the graph to look?"; "What does that tell you about the graph?" She asked these questions for the students to answer, gave sufficient wait time, and often re-worded a question if no one responded.

Ms. Henderson's GT lessons also incorporated the use of a graphing calculator. She had students use them as tools, but also made sure they used them appropriately and efficiently. Although students seemed very familiar with the calculators, occasionally questions specific to the limitations or capabilities of the technology came up during class. Ms. Henderson addressed these "side" topics, seeming to use them as

opportunities to reinforce concepts already learned. For instance, in graphing $y = 5x^2 - 100$, the class discussed the need for different scales on the x- and y-axes. Some students seemed unsettled by this, and she explained why it was perfectly all right to do this, although the graph would not look “normal.” Another time students were stumped when the calculator gave a vertex coordinate as .9999728, when they knew from the algebra that it should have been exactly 1. She explained that she believed the reason was due to the limitations of the pixel-screen, but admitted that she was not entirely sure.

Thus, Ms. Henderson conducted this GT class by consistently questioning students about concepts and relationships and having them question each other, formally having them work with partners or groups to promote further interaction, and integrating technology smoothly into the tasks at hand.

Ms. Henderson’s Instructional Practices

Tasks

In Ms. Henderson’s classes, the tasks in which students were engaged were always meaningful and relevant. The openers seemed to be carefully chosen so that they could be used both to check for understanding of prior material and to connect to the goal for the day. Ms. Henderson also clearly included a segment in her lessons that was reminiscent of Driver’s (1987) orientation/elicitation phase in which she helped to focus students on the task for the day, while drawing out their prior knowledge through a whole-class question and answer phase. She used motivational strategies that included hands-on tasks, focused thinking and writing time, demonstrations, and student-led

discussions. She consistently asked pointed questions to help students connect their knowledge. For instance, when a student initially chose the method of completing the square to solve a quadratic equation that in fact was factorable, she asked him, “When you finished, how did you know it was factorable [to begin with]?” Although calculator solutions were found and discussed in the GT class, they were never the main focus of the problem. They used the calculators more often to check their solutions or ideas, although certain peculiarities of the calculator were discussed along the way, such as scales and windows.

Many of the Pre-Algebra classes I observed included students doing geometry constructions. Ms. Henderson introduced students not only to the traditional compass and straightedge constructions, but also included an alternative way to produce the constructions using patty paper. This class seemed to be too large for her to adequately monitor what students managed to accomplish and understand from these constructions. However, as mentioned earlier, Ms. Henderson told me that her goal was only to get them practicing the physical maneuvering required in constructions, rather than to take away a rigorous understanding of why the constructions work.

Learning Environment

Similar to Ms. Drake, Ms. Henderson had administrative techniques and a teaching style that not only prevented classroom management problems, but fostered a cooperative atmosphere among the students. Certain supplies such as tape, staplers and a three-hole punch were permanently stationed on one side of the room for students’ use when needed; desks were arranged in a particular manner that allowed for much

interaction among students, but also allowed all students to have a clear view of the front of the room; students were put into groups on a permanent basis (but switched twice per quarter) so that “groupwork” or “work with your partner” came with concrete expectations. Class time also ran smoothly due in part to the way she organized the time. While students were working on the opener, other students would put homework problems on the board to get ready for discussion after the opener. The daily handout also provided students with a clearly written goal and expectation for the day.

What made the learning environment a setting for true meaningful learning was Ms. Henderson’s constant demand for interaction and justification of statements. Whether answers were wrong or right, she asked for justification and class reaction before rendering her own argument. Throughout homework discussions in both classes, she presented conceptual questions to students as if she really depended on their doing their homework in order to “be on the ball” in class. Students raised their hands eagerly, although she often called on students who had not been participating. In at least one instance when a wrong answer was given, she noted it, but praised it as a good “bouncing-off point” to help other students distinguish between two definitions that they had learned.

She often asked for students to give alternative solutions or methods in addition to those that had been mentioned and encouraged the students to question one another. She would validate their anxieties about difficult problems by saying, “I heard mixed reviews about the homework, so let’s discuss that.” She even had a 30-minute homework policy in all her classes, whereby students could stop after 30 minutes even

if it were not finished, as long as a parent signed off that they had spent 30 minutes working.

Ms. Henderson encouraged and expected students to think hard. Accordingly, she used appropriate wait times after questions, set aside time for them to orient themselves to problems on their own, and responded to questions with questions.

She also was not afraid of making a mistake, or letting her students express their feelings about fairness. For instance, in the Pre-Algebra class, after having students individually do a performance task from start to finish, she then set aside time not only to assess what they had done, but also to solicit their ideas about the task's directions, and its clarity or lack thereof. In the GT class, she initiated a discussion about whether a certain method was acceptable on a quiz, given the wording of a question. The discussion led to some important conceptual notions regarding the mathematics and resulted in her saying that she would make sure the directions were clearer in the future. On one occasion, I questioned her on something that she had taught that was not technically accurate, though not likely to cause any real problem for the students at any point. She reacted very graciously, saying that every year when she and another GT teacher read the particular relevant definition, they had trouble figuring out what their book was trying to say, and that my explanation finally made sense. The next day in class, she told the students that she had made an error, and that with my help, finally understood something that she had not before. She explained it to the students clearly and told them that they were expected to know the correct simplification.

A teacher is not always able to thoroughly address every point of a problem or discussion; instructional decisions must be made in a moment. As in the Pre-Algebra

class when there were certain elements of the mathematical content that might have been illuminated more explicitly, there were some isolated cases in the GT class as well. For example, during an opener problem, one group was asked to solve a quadratic equation by graphing. This group had some trouble. The teacher expected that they would graph using the methods they had learned a few weeks prior, such as finding the vertex, and then using the rate of change in the quadratic based on the leading coefficient. (She stated this after going through what the group actually had done.) In fact, the students used the intercepts, which was apparently acceptable to the teacher. Yet, using intercepts is really solving the quadratic in order to graph, rather than the other way around. Ms. Henderson, however, did not make this explicit, but instead chose to emphasize to the class the relationship between the x -intercepts and the solutions to the quadratic. She also did not make explicit the difference between the original quadratic *equation*, and the quadratic *function*, “ $y = \dots$ ” Perhaps this was already understood by the students based on their previous work with quadratics.

Discourse

As alluded to in other sections, Ms. Henderson was exceptional at engaging students in constructive discourse with her and the other students. Student-to-student interaction was formally and regularly expected through group-work and whole-class discussions. Whether answers were correct or not, and without giving any hint, she would say to the class, “Thumbs up if you agree, thumbs down if you disagree,” in response to a student explanation.

Ms. Henderson would also question the students constantly through every phase of a lesson, from the opener and homework discussions, throughout developing a new concept, to assessing understanding in closure. The questions went well beyond asking for an answer to a practice problem; they were almost always questions aimed at assessing students' understanding of the bigger picture. Even through topics in GT that were heavily procedural, she asked conceptual questions about the procedures, helping students reinforce their understanding of algebraic principles. There were occasions on which she might have pressed a student further to clarify a misconception before trying to address an incorrect response from her point of view.

Other themes

There were no other distinctive, consistent themes evident from the data on Ms. Henderson's teaching practices that need further elaboration. However, an interesting point of distinction between Ms. Henderson and the other teachers I observed concerned her attitude toward the state performance assessment; this was gleaned from observation, as well as a short conversation I had with her. During the period of my observations, I attended the school's eighth-grade "PA" simulation. Students came to school for only a half of a day and went to the room where they would be taking the State PA. I observed Ms. Henderson conduct this session with the students she would monitor during testing week. The students were not necessarily her own. After conducting an "icebreaker" group task, she spent about 15 minutes with the group eliciting their ideas about why it was important to do their best on the PA test and why results were important to the community and state. Her orientation was clear and

appeared to be sincere as she motivated the students, pointing out the kinds of skills that the test assessed. At the break, I asked her if she thought the state PA was a test for “gifted and talented” students only. She told me that on the contrary, she felt that the Pre-Algebra students actually had the advantage over the GT students because of the *content* on the test. (The students in GT were further removed from some of the geometry and statistics that the test might cover.) She was not concerned that these lower-track students might not possess the thinking ability required by the expectations the PA had, as Ms. Drake had insinuated.

Comparison of Ms. Henderson’s Pre-Algebra and GT classes

Similarly to the situation in Ms. Drake’s classes, Ms. Henderson’s GT class was much smaller than the Pre-Algebra class -- 16 compared to 27 students. Because Ms. Henderson questioned students constantly, there likely was less opportunity for each student to verbalize their own responses in a whole-class setting in the Pre-Algebra class, simply based on the class sizes. During the timeframe in which I observed, the Pre-Algebra students were also given many hands-on tasks that caused them some difficulty due to the physical dexterity required, and so those classes seemed a bit more chaotic than lessons I observed in GT. This could be accounted for by the nature of the tasks in the two classes being different, or by the sizes of the classes being more or less manageable, or both.

Homework discussions in GT generally lasted longer than those in Pre-Algebra. This difference could be accounted for by the mathematics content that was being developed in each course during my observations. In Pre-Algebra, much of the

homework that was assigned was self-checking in the sense of having a puzzle that students filled in as they obtained answers. The problems themselves dealt with definitions and relationships; they were not lengthy and procedure-ridden, nor were they application problems. Thus, in Pre-Algebra I only saw students explaining their homework solutions on the board the first day I observed, which was the end of an algebra topic. In the GT class, however, they were in the midst of an algebraic topic that required some cumbersome manipulations, and that included application problems requiring many steps, including the difficult task of making sense out of the context of the application problem in order to write an equation.

Ms. Henderson used questioning in both classes similarly. That is, she questioned students about problems whether students expressed difficulty or not, and she questioned them about the concepts embedded in the mathematics at hand. In the GT class, questions were focused on making sense of the algebraic procedures that the students were learning; in Pre-Algebra questions were focused on the connections between the geometric properties that the students were learning. Again, perhaps due to the nature of the mathematical content, the questioning in GT seemed to have had more depth than in the Pre-Algebra, where much of the new content was definitional.

There were more cases in which a student in the GT class initiated a discussion based on an idea they had about the mathematics under investigation. This allowed me to observe Ms. Henderson's handling of situations in which flexibility would be required to appropriately respond to an unexpected situation. Ms. Henderson responded to students' suggestions, questions, and ideas in a way that encouraged ingenuity and interaction. There were fewer such diversions evident in Pre-Algebra, because there was

less student-initiated discussion. This could be due to the larger class size or to the students themselves. In the GT class, it was almost always the same one or two people who vocalized an original idea. In general, Ms. Henderson appeared to treat the two classes similarly in terms of expectations for students to think through concepts, to question each other, and to respond to questions.

PA Alignment with Instruction and Opportunity to Learn

The PA had very transparent expectations about students being able to defend and justify their responses. Through her questioning techniques, Ms. Henderson required her students to anticipate outcomes, make connections, reason logically and justify answers. Students seemed very involved in class discussions and eager to respond. In each class, on several occasions, students were required to write out a response to an open-ended question about a mathematical procedure or concept or to write out how they would proceed toward a solution. Students were certainly used to having to provide a reason for their thinking, and this was commensurate with the expectations on the PA.

In both classes, students were given tasks with written directions. Ms. Henderson always reminded them to read through the directions circling the verbs that told them what to do and underlining key information. In this way, she was developing their facility to read through and make sense out of a task. On one occasion in the Pre-Algebra class, she actually had them simulate a testing situation, by giving them a set amount of time to carry out a hands-on task individually. She spent a lot of time afterwards discussing their understanding of the directions and where some students had

gone wrong. The task itself was an extension of concepts they had studied the previous day. For the most part, the tasks Ms. Henderson provided fit the NCTM's description of "worthwhile mathematical tasks."

Another expectation on the PA is group work. Some questions on the PA are actually begun with students working in groups. Ms. Henderson had formal pairs and groups established in both classes and made use of these various formats often. In the GT class, she even gave an in-class pairs-quiz, although the students had to write up individual responses following the time they were given to work in pairs. In both classes, I observed students turning to their partners for clarification and direction often. This kind of interaction seemed very much encouraged.

Ms. Miller

Ms. Miller was teaching in a school with nearly 1000 students. The school served a predominantly working class area and 42.4% of the students at the school received free or reduced lunch. The minority enrollment at the school was almost 28%. Ms. Miller taught two eighth-grade classes of GT (Algebra II) and three eighth-grade classes of Algebra I. I observed her teaching a GT class and then an Algebra I class, back-to-back in the middle of the school day. These courses were the middle- and high-level courses for eighth grade at this middle school. Because there was no time in between these classes, any conversations we had generally took place after the second class, which was her lunch period.

Ms. Miller had her room arranged in the manner shown in Figure 6, which allowed students to focus on the front of the class, but still have some opportunity to

interact with one another. The desks were arranged as shown for both classes, but the circles depict where the 25 students were seated for the Algebra I class.

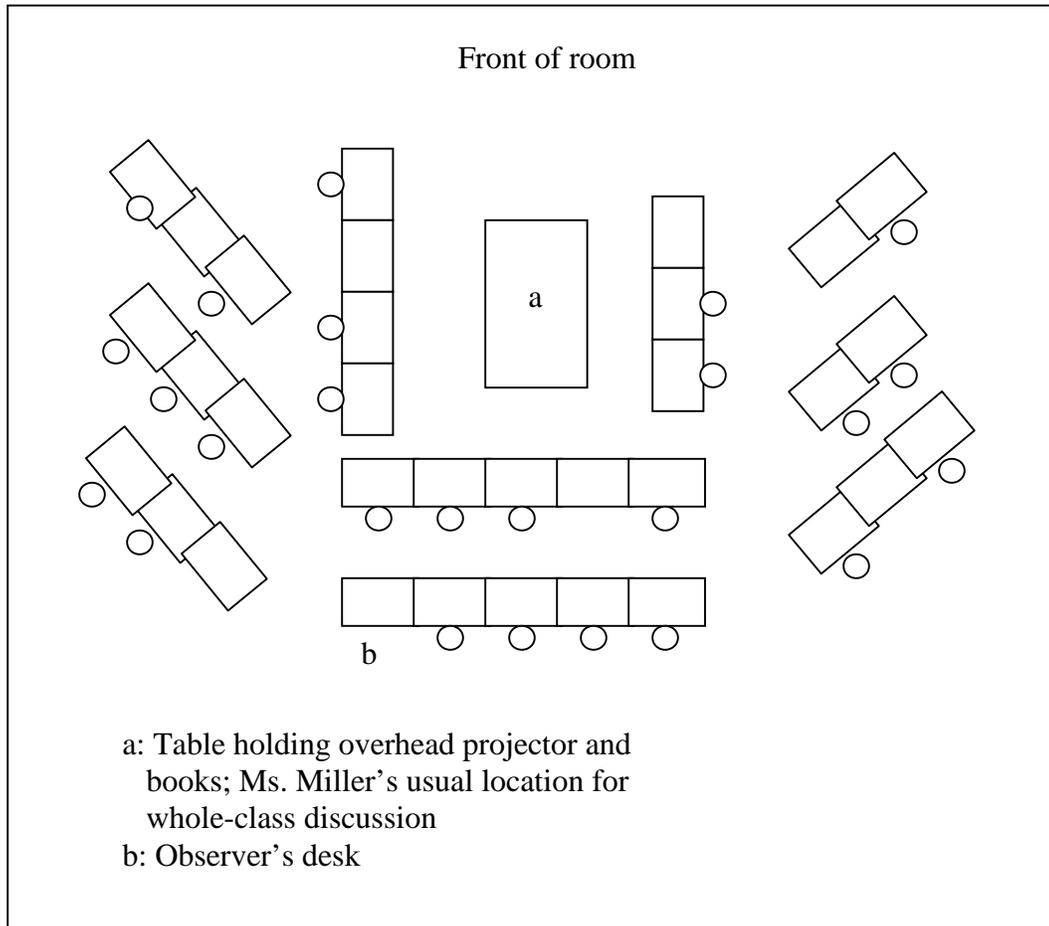


Figure 6. Arrangement of desks and students in Ms. Miller's Algebra I class.

Algebra I

The Algebra 1 class had 25 students. I observed the class 13 times over a five-week span, although the middle week was the district's scheduled spring break. Table 23 shows the timeframe over which I observed, and whether an observation took place or the reason one did not.

Table 23

Time Table for Observations of Ms. Miller's Algebra I Class

| Date | Observation / No observation |
|---------------------------------------|---|
| Monday, April 3 | Observation |
| Tuesday, April 4 | Observation |
| Wednesday, April 5 | Observation |
| Thursday, April 6 | Observation – full period test given |
| Friday, April 7 | No Observation – I was unable to attend |
| Monday, April 10 | Observation |
| Tuesday, April 11 | No Observation – full period re-test; I left after test was given out |
| Wednesday, April 12 | Observation |
| Thursday, April 13 | Observation |
| Friday, April 14 | Observation |
| Monday, April 17- Monday, April 24 | No Observations - spring break |
| Tuesday, April 25 | Observation |
| Wednesday, April 26 | Observation |
| Thursday, April 27 | No Observation – “Take Your Kids to Work Day”; classes effectively “canceled” |
| Friday, April 28 | No Observation – Field Trip; classes canceled |
| Monday, May 1 | Observation |
| Tuesday, May 2 | Observation |
| Wednesday, May 3 | Observation |

During my observations, Ms. Miller's Algebra I class had just finished a chapter on polynomials and completed their unit test on that topic. Due to poor performance overall, the class was given a re-test on this unit three class-periods after the first test.

The class also began and nearly finished a unit on probability during my observation period. After the introductory lesson, which included the definitions of and practice with theoretical and experimental probability using dice and coins, students spent the rest of the unit setting up and carrying out probability simulations that utilized real-life contexts.

The division of the class periods into regular, distinct segments did not appear to be as customary in Ms. Miller's Algebra class as they were in the lessons of the other observed teachers. As with the other district classes, there was always an initial "drill" exercise, which the students did individually and then went over as a whole group. During the first week that I observed, Ms. Miller completed the last chapter in the district curriculum that would define topics for the final exam. During this time, I observed Ms. Miller going over homework from the text and lecturing on new algebraic content, following the drill exercises.

Particularly during the weeks following the unit test on polynomials, after the drill, class time was spent in varied ways with many days having what seemed like quite a bit of spontaneous review. This could be because it was nearing the end of the year. Beginning with the second week of my observations, Ms. Miller began giving out review packets on Mondays that were due on Fridays; these were intended to prepare the students for their final exam. The remaining class time was spent on the topic at hand, which was probability. Generally, when Ms. Miller began a new simulation activity, she used what remained of class time to introduce, or "analyze," the activity and get the students to a point where they could carry out the simulation activity on the following day. Homework during these weeks was minimal, or at least minimally

discussed. The homework was merely answering the questions on the simulation activities that focused on summarizing and interpreting the data collected during the in-class simulations.

Below, I describe first the drill exercises, since these occurred on nearly every observation day. I also describe briefly the time used for review. In describing drill and review sessions, I discuss some of the interaction that I observed, and how Ms. Miller handled algebra content. Then I describe in detail the instruction that addressed the content for the end of the polynomials chapter, since it illustrates how Ms. Miller introduced an important topic that has some fundamental algebraic concepts embedded. Finally, I characterize in more general terms how Ms. Miller carried out the probability unit using the simulation activities.

Lessons began with two drill exercises on the overhead projector as students walked into the classroom. These exercises were almost always unrelated to the current or recent days' topic, but rather were general algebra review questions. On only one occasion did the drill serve to get students ready for the activity for that day, by having students read the paragraph written on the activity handout and answer a question that required them to repeat what was written in the paragraph. Ordinarily, students worked individually to complete the drill problems. At some point students would volunteer to put their work on the board. Then, as a whole group, the class would discuss the problems. All together, the drill segments I observed lasted from 6 to 15 minutes. The way in which Ms. Miller conducted the whole-class drill time varied. That is, sometimes she would refer to a solution put on the board by a student and ask the class, "Agree or disagree?" or "How do you like it?" This occurred on occasions when the

problems were done both correctly and incorrectly. However, student responses were minimal. This could be due to short wait time or to the fact that students knew that she would soon express her own opinion rather quickly. For instance, a student put his work on the board for a problem in which an expression had to be evaluated by using the order of operations. There were several mistakes in the solution. Just as a student began to respond to her question about “how [the class] likes it,” Ms. Miller erased the board, and asked the whole class, “What concept should we be following here?” She led the students to recall the pneumonic device they had learned for the order of operations. On most occasions, Ms. Miller talked through the student-solution herself, fixing any major errors as she went. As with homework discussions and review packet questions, on occasion, she had the students, themselves, talk through their work. However, she seemed not to be consistent in this aspect of classroom interaction.

In eight consecutive class periods, there were six days wherein the class spent between 11 and 20 minutes either asking questions from a review packet, going over review packets handed back, or reviewing a test that was given back. The review packets were summative assessments that the district had just developed and had given to schools without specific direction about how they were to be used for this academic year. Ms. Miller allowed students to work on their review packets on three observation days, usually lasting as much as 20 minutes. During this time she took questions from students either individually or in a whole-class setting. When she discussed work that was handed back, generally the class focused its attention on Ms. Miller while she went through the common errors she came across.

During the sessions when she let students ask her questions regarding the review packets, she generally responded with other questions that would lead the student through a procedure. For example, a student asked about a problem that said, “What is the slope of the line represented by the table of values [below]?” Ms. Miller asked the student the following questions, one after the other, as the student gave her the correct responses: “Do you know how to find slope?”; “What are those things they gave you in the table?”; “How many points did they give you?”; “How many do you need to find slope?” This kind of questioning occurred fairly frequently, and although the questioning manner of explanation was procedural in nature, it did appear to help the students through the problems, and at least momentarily help them to model the thought process for the kinds of problems on district algebra exams. However, if a student did not answer a question that she posed to help them answer their own question, she was fairly quick to respond to the question herself. For example, a student asked about a problem in which they had to choose one of four graphs that would match a given linear inequality. He asked, “How do you figure out the shading?” She responded, “What does the sign need to be since the shading is below?” When she didn’t get a response, she continued, “Less than is below and greater than is above.”

Ms. Miller’s explanation illustrated another key characteristic of her explanations; they were often too general, and therefore incorrect. That is, in this case, her response indicated that students likely encountered only problems in which “ y ” was isolated on the left side of the inequality, in which case her explanation is correct. However, if “ y ” is not isolated as such, this is not correct. Thus, in the general case, her explanation was incorrect. Another example of an inaccurate explanation was when she

was helping the students through a problem involving the Pythagorean Theorem, written as $a^2 + b^2 = c^2$. She asked, “What is c ?” The class responded by saying the hypotenuse. She asked, “And the hypotenuse is always...?” She then answered the question herself, saying, “The one we look for on the diagonal.” This explanation assumed that triangles would always be oriented in a particular way. Such rule-bound explanations are likely to lead to content misconceptions for the students.

Ms. Miller often asked clarification questions on algebra content, even when correct answers were given. For example, when a student correctly said that $(2x)^3 = 8x^3$, Ms. Miller asked the class, “Why is [the coefficient] not six?” However, she often used and accepted imprecise language, as in the case when she explained that when a student correctly carried out the multiplication in $4x^3(2x^6 + 3x^2 - x + 1)$, that student “just multiplied the numbers and added the variables.” In another case when a student correctly wrote that $.0000062 = 6.2 \times 10^{-6}$, Ms. Miller asked the class, “Why is she using a negative?” When a student responded, “Because we’re going backwards,” Ms. Miller simply replied, “Good.”

The following description addresses the initial lesson on solving quadratic equations by factoring, which occurred during my first observation. That day, together with the succeeding lessons, demonstrated significant characteristics of Ms. Miller’s teaching. The block quotes are edited directly from my narrative field notes written immediately following the observation. My interpretive comments are written in parentheses.

The second drill question was: “A triangular block has a height h and a base of $4h + 6$. Its area is 8. Find the value of h .” Ms. Miller solicited responses from students to draw a picture of a triangle with base, $4h + 6$ and height, h . Then she said, “What do I do now? Jimmy started to write $a^2 + b^2 = c^2$. Why can I not use

this?” There was a pause. “What kind of triangle do I need for this?” Someone said, “Right.”

Next Ms. Miller said, “There’s something we need to use and there’s a key word in there.” She led them to the area formula, reminding them that if they forgot it, they should have looked it up in their notebooks. She wrote on the board: $\frac{1}{2}bh = A$. Then she solicited student responses to use substitution to obtain: $\frac{1}{2}(4h + 6)h = A$. Then a student told her to distribute and he said, as she wrote: $(2h + 3)h = 8$. Ms. Miller substituted the 8 for A; the student had not told her to do so. After asking students what to do next and not receiving much response, students reminded her that they “didn’t get that far.” (I thought perhaps she had been absent on Friday, but after speaking with her after class, she just said that she teaches three of these classes and forgot that this class had gotten “stuck” on the factoring, so that they hadn’t gotten to solving equations.)

Once she realized that they hadn’t solved quadratics, she apologized to the class, and it appeared as if she were about to begin something else, but someone asked her to finish the problem anyway. Thus, by way of introduction, she said, “We will solve this equation using GCF or binomial factoring. We can break up each part of the equation and set it equal to the solution, 8.” (Ms. Miller field notes / post-observation reflections, April 3, 2000)

Her explanation erroneously applied a procedure that makes use of a fundamental concept, called the zero-product. Thus, this introduction to solving quadratic equations was incorrect. Ms. Miller did not mention the zero-product concept; she also did not refer to the “parts” as factors.

Ms. Miller continued her solution on the chalkboard, where she set each of the factors equal to eight. She completed the problem, having obtained two “solutions” for the height, h , but did not give any explanation for the answer or the fact that there were two answers. She also did not check her solutions, even though she later included “checking” as one of the steps in solving quadratic equations.

Next, she appeared to begin the formal lecture segment of this class. She wrote the title “Solving Quadratic Equations by Factoring” on the overhead projector. She also listed the numbers one through four vertically, apparently so that the class would

fill in the steps required to solve a quadratic equation as they learned them during the lecture. She wrote the first example on the overhead: Ex.1 $x^2 - 11x + 24$

Her plan seemed to be to talk through the example and write in the steps as they proceeded through the solution. Her first question was “If I was to give you a trinomial, is this an equation or an expression?” Several students said “expression.” She said, “I want to turn this into an equation. What can I set it equal to so as not to change its value?” She continued, “Zero, since you can add or subtract to zero and nothing changes.” (Ms. Miller field notes / post-observation reflections, April 3, 2000)

Ms. Miller’s question did not make sense. That is, the value of an expression depends on the value of the variable. When an open expression is set equal to any particular value, the expression is no longer “open.” That is, the value of the variable is then determined. Furthermore, she didn’t add or subtract zero; she set the expression equal to zero.

Next she wrote: $x^2 - 11x + 24 = 0$ and asked the class, “What’s your first instinct?” (At that point, I’m not sure if a student said factoring or she did. Thus far, she had not made any mention of zero-products.)

She continued writing on the overhead: $(x - 8)(x - 3) = 0$ and asked, “What was the next thing I did on your drill?” She said, “I took each piece and broke them apart and...set them to the same thing.” (Note Ms. Miller’s imprecise mathematical language.) A student said, “You didn’t set the other one equal to 0!” to which she replied almost impatiently “because it was equal to eight!” (Ms. Miller field notes / post-observation reflections, April 3, 2000)

At this point Ms. Miller had filled in some of the steps, as (1) Factor quadratic; (2) Set each binomial equal to 0; and (3) Solve for x. However, note that step 2 was not consistent with her explanation to the student nor was it consistent with her solution to the drill problem.

As Ms. Miller was preparing to obtain the solutions to the problem, $(x-8)(x-3) = 0$ was written on the overhead, and she said to the class, “I want you to be mature enough to visualize what your answers are without writing

anything more down.” (Apparently, she wanted them to notice that $x - 8 = 0$ implies that $x = 8$, and so forth.)

Students offered a number of wrong answers. One student suggested $x = 7$ and began to explain, “Because there’s a 1 where x is...” Ms. Miller dismissed this and said, “No way!” Some other kids offered the correct answers of $x = 8$ and $x = 3$. Ms. Miller seemed to want the students to see why the answers were 8 and 3 without writing out the solution to the linear equations. Someone else noticed that the signs of the solutions were the opposite of the term in the binomial, and Ms. Miller reiterated aloud “If it’s a positive, it’s a negative. If it’s a negative, it’s a positive.” She did not ask the class why this was the case. (Students might be confused when they encounter a factor such as $(-8-x)$. The “rule” she gave them does not work in general.)

She then said, “Every time we solve an equation, we should check our work!” However, she only checked the solution $x = 3$ and did so without writing the original equation. This made it difficult to follow what she was checking. She wrote only: $9 - 33 + 24$; $-24 + 24$

Ms. Miller then asked, “Did it check?” She wrote: $x = 3, 8$ and said, “Those are called roots because if we made an ordered pair, we’d get $(8, 0)$ and $(3, 0)$ and ...” She continued, but not as loudly. (She appeared to be unsure of herself momentarily.) She said that those would be zeros on the x -axis if they graphed. By that point she had filled in all of her steps, including the fourth step, which was “Check your solution.”

Ms. Miller then did another example of a quadratic equation with zero on the right side. She factored the trinomial for the students and then wrote the answers, without writing the two equations. She said, “If you don’t understand where these [roots] came from, you need to write down and solve: $x - 6 = 0$ $x - 3 = 0$.” Some kids were saying that they were lost. Another student said, “So just switch signs!” She replied, “At this point, yes.”

About ten minutes before the period ended, she had the class refer to their books as they went through two problems aloud. She said, “#2 - Raise your hand if you can find two roots of x without writing anything down... Those of you not raising your hand, is there something you’d like to ask or see?” No one said anything. Next she asked them another problem from the book. The problem read: “True/False:

If $(x+3)(x-3)=0$, is it true that $x + 3 = 0$ or $x - 3 = 0$?” A student gave the answer, “true.” There was no further discussion about this problem. I noticed that she did not ask the class to answer the previous problem in the book, which asked a similar question except that the right side of the equation was one, rather than zero. Those two questions together would have offered a good opportunity to highlight the misconception presented to the class during the drill.

Next, Ms. Miller asked for the solutions to the problem: $(x + 1)(x + 2) = 0$. After obtaining the answers to this, she explained briefly why that equation implied that $x = -1$ by reminding the students that when we subtract the 1 from both sides of the equation $x + 1 = 0$, the result would be $x = -1$. However, she only stated this orally; she did not write down the distinct linear equations.

Ms. Miller then said, “Everyone is on board.” She asked, “What if we now have a coefficient other than 1?” and wrote on the overhead:

$$2x^2 + 5x + 3 = 0. \text{ “Can we still factor?”}$$

She walked around the room to check students’ work. Some students were still having trouble factoring. She then went to the overhead and wrote:

$$(2x + 3)(x+1)=0$$

$$2x + 3 = 0 \quad x+1=0$$

Pointing to the first factor, she asked, “Why can’t most of us automatically give an answer?” She said, “I don’t want to show my work. What would I write?” A student said aloud: “ $2x = -3$; $x = -3/2$ ” (She seemed pleased that the student was able to obtain the answer without writing it.)

She ended the class by saying, “So you’ve learned two things today. When the coefficient is 1 and when it’s something other than 1.” She did not summarize the idea that they had just solved quadratic equations, which were new kinds of equations for them.

She assigned homework problems from the book and students put their books together waiting to be dismissed. Once the students had quieted down and were ready to be dismissed, she asked them for the four steps the class had developed while she asked them leading questions. As each group answered, she allowed that group to leave. (Ms. Miller field notes / post-observation reflections, April 3, 2000)

While Ms. Miller was making those serious mathematics content mistakes, I was battling over whether I should say something to her, especially because it was my first day observing. The drill problem with the triangle was still on the board at the end of the period. I had decided that I needed to gently remind her about the zero-product rule, as I had thought she should bring this mistake up with the class the following day and address the concept. Thus, after class I approached her and thanked her for letting me observe. I said that I wanted to tell her something I noticed was a mistake, but that I was aware that she had been caught off-guard when she realized the class hadn't covered this. I also told her that I understood that sometimes it is difficult to notice one's own mistakes when standing right in front of the board, but that from the back of the room it was more noticeable. I was trying to let her know that this was not a big deal, and I ordinarily wouldn't do this. I did not want to make her feel uncomfortable with my observations. Pointing to the factors set equal to eight, I told her that one couldn't actually do that. I said, "With zero, you're saying that 'this' times 'that' equals zero, so one of the factors must be zero, but when 'this' times 'that' is eight, the factors can be four and two or one and eight, or..." She finished my sentence when she said the factors could be anything. I mentioned the concept of zero-product, but I'm not sure if she understood the importance. She said that when she had done the problem earlier with other classes, that problem came out to be a prime trinomial -- it couldn't be factored, so she just wanted the students to get to "set it up."

The next day, after drill problems, the class went over the homework from the previous night, which consisted of equations that had quadratic expressions on the left side and zero on the right side. Ms. Miller reminded the class of the steps given the

previous day with no mention of the mistake she had made. She also did not mention the check step, and in fact none of the homework problems were checked, none of the problems in this class period got checked, nor was checking ever mentioned. When she discussed the first homework problem on the board she said, as if reminding the class, “We can solve only when?...it’s equal to zero,” even though this statement contradicted the drill problem from the day before. As they finished discussing the homework problems at the board, Ms. Miller also reiterated a rule she had given them the previous day by saying, “Whatever sign is in the parentheses, the solution is always the opposite – always.”

After the homework discussion, Ms. Miller said to the class that they would go back and pick up where they had left off the previous day. For the next 16 minutes of the period, Ms. Miller wrote a series of quadratic equations on the board and solved them using a standard factoring technique. She continued to be insistent that students not actually show any steps after the factoring step, other than the solutions. She told them that they should be able to “visualize” the solutions. Throughout the examples, she did not mention the zero-product rule and did not communicate any reasoning involved in the procedures; the problems were presented in a purely mechanical fashion. Some of the examples that she presented required factoring on the left side; others were already factored. However, there were no problems given in which the left side was already factored and the right side was something other than zero, so that she was not faced with precisely the same situation as she had been in the drill problem the previous day.

After 16 minutes of examples, Ms. Miller gave the students two problems to try on their own, as she went around the room to observe and help. She announced several

times that they were very far behind, which is why she needed to give them a double dose of homework; she said that she wanted them to be ready for the test on Thursday.

The next day the class spent some time reviewing the entire chapter to get ready for the test, and then Ms. Miller conducted a whole class lesson in which they solved an application problem that made use of solving a quadratic equation by factoring. A question very similar to that one was on the test, which they took the next day. The test itself was one of the district's summative assessment's that had just been given to the schools, with no direction about how they should use them that year. It was formatted similarly to the final exam in that it had multiple choice problems followed by constructed-response items, most of which were context-bound application problems. This test seemed to have high standards, but based on what I observed her teaching and reviewing, I questioned how fair the test was. She told me after the test period that the test didn't exactly match the curriculum, so that she'd likely have to eliminate some of the problems. However, students were not told about this ahead of time and could have become frustrated and discouraged while taking the test.

On two separate occasions Ms. Miller talked to me about the unit on probability that she was doing. The first time, she showed me an activity book on probability and statistics that she said she thought she would use as supplements that year. She said she did not like what her book offered; she wanted to do "some hands-on stuff." On a subsequent occasion, after she had mentioned something to the class about having finished the curriculum, I asked her if the entire probability unit was her choice as extra material. She told me that no, the district wanted her to do simulations, but she wanted

to start with easier ones than the book had. She said, “They’d never be able to do the ones where they have things like ‘there’s a .7 chance of x and .3 chance of y.’”

There were ten lessons on probability during my observation period. The first two lessons were conducted in between the test on polynomials and the re-test on that topic. The other eight lessons occurred on consecutive class periods, although school was interrupted during that time by spring break, “Bring Your Kid to Work Day,” and a field trip. After providing some detail about the first two lessons, which apparently served as introductory lessons, I will broadly describe how the time was spent during the four simulation activities and what the students likely did and did not take away from the unit.

Unfortunately, I was unable to attend the class period when probability was first introduced to the students. I did, however, obtain the handout that was used, and talked to Ms. Miller briefly about what the class had done that day. On that day, students were given a three-page handout. The first page had basic vocabulary words that the class went over as a whole group. They filled in definitions for probability, theoretical probability and experimental probability. The second page of the handout described a “carnival” that had six game booths. Each booth had a game of chance that was explained briefly. For example, the directions for the “Pick a Number Booth” said, “Predict what number you will roll. Then roll a number cube. A true prediction scores one point.” Some games were slightly more complex; the “Coin and Cube Booth” said, “Toss a coin and roll a number cube. Tails and 3, 4, 5, or 6 scores a point.” The third handout was a table that listed the six games vertically, and asked the following questions across the top: number of ways to score; number of possible outcomes;

theoretical probability of scoring in 1 turn; and expected number of points in 10 turns. Ms. Miller told me that they spent the class time figuring out the number of possible outcomes for each game. I was not sure if they discussed the expected number of points in 10 turns.

The next day, which I did observe, students were given 14 minutes to work with a partner, whereby each partner was to play any 10 carnival games he or she wanted. They needed to record the points they earned. After students noisily played the games, Ms. Miller told them each to go up to the board and put a tally mark next to the number of points they earned. They were not asked to record the specific game they played. After everyone copied the tally chart off of the board, she asked them to copy down the three homework questions. These questions were 1) Make a graph of the class results; 2) How did your points compare with the totals of others in the class; and 3) Which booths would you choose to play and why? After students did this, the remaining 21 minutes of class was spent going over the test they had taken the week before.

There was no discussion of the results of the “carnival” whatsoever during this class time. A discussion could have served to bring out or reinforce a number of important concepts in probability. For instance, the third homework question might have revealed students’ basic understanding of probability. However the activity was not executed, and the data was not recorded, in a manner suited to answer this question. What they needed to do was keep track of points won for the different games. The first two homework questions seem unrelated to the concepts one would think are central to this introductory activity, and indeed to the data they individually collected. When the class did go over the questions two days later (since the day following this activity was

devoted to the re-test for the polynomial chapter) none of the substantive portions of the activity were discussed. Ms. Miller talked about the graphs some students turned in, mentioning the PA exam, and how important titles and labels of graphs were. Then she read aloud some of the sentences written to answer question 2. She was rather angry with the class, and was merely making the point that their sentences were so bad, it was hard to decipher what they were saying. She said, "I feel bad for Ms. T. I don't often have to read your stuff, but she does." She then went over the third question, which did ask a significant question related to the activity. She read an answer aloud that was nearly perfect, but very briefly pointed out that the student had nicely answered which booths were more likely to yield winners and why. It was not clear whether other students understood the significance of the question and the answer, as no follow-up questions were asked.

The entire discussion of this homework took only nine minutes. There was no interaction between students to ask them what they had chosen to play, how they did, and whether they were surprised at results. If students did not understand the concepts or the point of the activity when they completed it, no additional opportunity to do so was possible during this discussion time.

Following the carnival activity, which was intended to acquaint students with using coins and dice to play games of chance, Ms. Miller gave out a series of context problems whose main question was a probability problem. The class was to set up experimental simulations using coins, dice, spinners, or random number generators to carry out the simulation and ultimately answer the questions.

The activities, themselves, were challenging in terms of asking students to think through how to simulate various experiments. However, in most cases, Ms. Miller did not give students the chance to come up with the actual simulation on their own. She generally took control over setting up how to conduct the simulation, and even closely directed the students while they carried it out. Each lesson began with Ms. Miller giving out the activity and asking students to read it silently. She generally asked some questions to make sure they understood the situation described. In some cases, the activity was worded rather vaguely and so had the potential to be interpreted differently than what was intended. Usually before the end of a class period, the students were at the point where they were ready to start the simulation. She spent much of the introduction phase getting students to choose an appropriate simulation tool. At the end of the introduction to the first of these activities, she asked students, “What tool we choose depends on what?” She answered along with some other students, “The amount of choices we have.” This answer was not entirely correct; she was ignoring the issue of equal or unequal outcomes.

Ms. Miller made other mathematics content mistakes during these activities. One such mistake is described for the first simulation, which was “Dutch Elm Disease.” The gist of the problem was to figure out how many trees of six would remain healthy after four months of exposure to disease. The activity read:

Suppose there is a stand of seven elm trees in an area near your home. The bark beetle is in a feeding stage during the months of June through September. A diseased elm tree is surrounded by healthy trees. Each healthy tree has a 1 out of 6 chance of becoming infected.

The reading should have more clearly emphasized the fact that the trees had a 1 out of 6 chance of becoming infected *each month*. The first day they received this

handout, the class merely read through the problem and discussed tools to use for simulations. Below are some edited notes I wrote following the lesson in which they carried out the simulation:

In Algebra I today a probability simulation ended up being focused primarily on rolling dice. Students were not given any time to discuss or think about the simulation they were carrying out. In fact, it was done incorrectly.

Ms. Miller began the simulation part by asking the students what tool they would choose to carry out the simulation. Although she said that a spinner and random-number generator could also be used, they would use a die. She asked them, "We'll roll a die to find out what?" (I think the answer that someone gave was to see if a tree got infected, but as described below, what the class actually did was use it to see *which* tree became infected.) Ms. Miller quickly explained, "For June, we have to roll the die 6 times, once for each tree." (What she needed to do, was to make sure each roll was for a specific tree and that there was a $1/6$ chance it would be infected. In fact, what she ended up doing was using the number that came up as a particular tree that got infected.)

She rolled 6 times, and all numbers that came up were infected trees. If a number came up twice, it just meant that that tree was already infected. So for June, on trial one, the class got 4 distinct numbers, so that four trees were infected. (Doing it this way, in June, any one tree had a $1 - (5/6)^6$, or .665 chance of being infected. This was far different from $1/6$ or .167.)

Students still seemed confused, so she reiterated, "Every tree has a $1/6$ chance of being infected so we have to roll 6 times." In subsequent months, they only rolled the number of times that there were trees left, looking to see if those trees came up. (These trees had a different chance of being infected on subsequent months.)

The result was that the class ended up with nearly all trees infected at the end of each trial. (In fact the expected number of infected trees at the end of 4 months should have been less than 4.)

At the end of the ten trials which were done by the whole class together with a lot of commotion, Ms. Miller asked the students to answer question E on the sheet, "Based on the results of your simulation, what is the probability that at least 50% of the trees in the stand will become infected with the disease?" Students offered: $1/10$; $10/10$; $5/10$. She did not address these answers, but asked, "What does the last number [in the last column] have to be for 'at least 50%'?" They said 3. She asked how many were at least 3? They answered 10 out of 10. She said, "so 100%." This entire exchange took less than 1 minute. There was no discussion. She asked them to answer questions F and G for homework, and quickly discussed writing a letter. (F asked, "How many trials should be done to insure that your results are accurate?" G asked them to write a letter explaining their opinion as a community member on whether money should be spent to protect the trees.) (Ms. Miller field notes/post-observation reflections, April 13, 2000)

This activity should have offered much opportunity for discussion about several concepts in simulations, such as the importance of repeated trials. However, Ms. Miller did not execute the lesson as it was likely intended. By negotiating the activity down to the level she did, students merely had an exercise in rolling a die. Determining how best to simulate this experiment was the challenging part of the problem. The teacher did not give the students any chance to try to think about it themselves. She and they were confused about the numbers on the dice and the names of the trees. They might have discussed using a spinner with six spaces but one labeled “diseased,” or a blank die with only one side labeled “diseased.” Students did not have a chance to think through this problem, as Ms. Miller took strong command. Yet she made rather significant content errors fairly often. This exercise addressed the State PA content and process, as it required students to write and defend an opinion at the end. However, because important concepts were not discussed, it is not clear that students would be able to write adequately about the topic.

After class the next day, I told Ms. Miller that the probability experiment from yesterday had been “driving me crazy on the way home.” I said I couldn’t figure out what was bothering me. She said she thought something was wrong with it -- that they discussed it in math department, but when she went to another teacher to show her what she was going to do, that teacher said it was right. I then told her what was wrong with their simulation. She said that she didn’t think they were supposed to end up with near 100% infected trees on the final question, but didn’t know why they were getting that. She seemed to understand what I told her, but as before with factoring, she didn’t seem

too worried about the effect on the students. In fact, she didn't say anything about wanting to address the error with the students, and indeed never did.

The other simulations, although technically carried out correctly, also lacked adequate discussions on the main points. For instance, the class never actually discussed the *use* of simulation in answering the main question on each activity. Ms. Miller simply began the activities by essentially stating they would need to conduct an experiment. By the third simulation, students were able to offer ideas when she asked them probing questions to get them thinking about how to do the experiment. However, wrong ideas were ignored or dismissed, and there was only limited discussion of good ideas. However, once Ms. Miller did call a student to the board who had given the idea for a chart and had him conduct a whole-class discussion to set up an effective chart for the simulation. Although this part of the lesson seemed particularly fruitful since many students helped set up the chart, I noticed that at the bottom of the handout there was a hint that explained how the simulation could be carried out. No one in the class ever acknowledged that paragraph, although it explained precisely what students offered aloud.

The third simulation addressed whether water would be able to flow from point A to B, depending on whether two pumping stations would be open. Between points A and B there were four different possible routes for the water to take, each route passing through two pumping stations. Unfortunately, when the class was ready to carry out the simulation the following day, Ms. Miller took control over how the class should do it, and had them try to carry out 50 trials as a whole class. This ended up taking more than 35 minutes over two class periods because of the chaos involved with having five

different students flip a coin (simulating five pumping stations) and relay the result to the whole class for each trial. No one suggested breaking into groups and combining data, even though that was suggested in a hint provided on the bottom of the activity handout.

Again, discussion of the important concepts was limited. They never discussed why the activity suggested 50 trials or the law of large numbers, which was a major concept embedded in this and some of the other simulations. The following episode illustrates Ms. Miller's instructional decision-making.

After the first day of conducting the pumping station simulation, students had to complete the first 10 trials and then answer the question, "On average, how many stations were working at any time, according to your simulation?" When Ms. Miller began going over this question the next day, she asked, "How many stations are there?" The students responded by saying, "five," which was correct, but not what Ms. Miller intended. She told them that a station was when two stations in a row are working. (This was actually a *path* of pumping stations, and it was slightly more difficult to identify the number of these working on any one trial by looking at the recorded data.) The students did not disagree; rather they started giving her the number of pairs (paths) working in trial order. On the fifth trial, she heard a student say "five." As Ms. Miller started to say that it couldn't be five, she looked at me for verification and I mouthed to her that each of the five was its own pumping station. She turned to the student who had answered five, and said, "OK, you're right" and continued on, now having changed her definition. She did not explain to the class that she switched definitions. Apparently some of the students followed this line of reasoning and gave her the correct numbers

for the rest of the remaining trials. She did not go back and address the first four responses, which were wrong. After listing the ten numbers on the board, she found their average. There was no further discussion of this.

At the end of the period, after the class as a whole had finished all 50 trials, the students were told that their homework was to go back and answer all three questions again, for all 50 trials. One student asked about the averaging question. Ms. Miller pointed to the ten numbers still on the board from the earlier discussion and explained where she had gotten them. The student noticed that some were wrong, and Ms. Miller said, “Oh. Something must have happened.” Ms. Miller’s initial mistake seemed to be a temporary misunderstanding of the activity’s description of a pumping station. It seemed that she was fully aware of this mistake when she changed in mid-stream and began recording a different kind of number. Yet she did not clarify this adjustment to her class.

Algebra II (GT)

The eighth-grade GT class I observed had only 11 students in it. The other eighth-grade GT class had 15 students, and was also taught by Ms. Miller. The desks were arranged as shown in Figure 6 for the Algebra I class, and since there were only 11 students, many seats were empty. Students sat spread apart, with at least one empty seat in between any two students. I do not know if students or Ms. Miller chose this arrangement, but it seemed as though students had the choice to sit where they wanted. Table 24 shows the timeframe over which I observed.

Table 24

Time Table for Observations of Ms. Miller's Algebra II (GT) class

| Date | Observation / No observation |
|---------------------------------------|---|
| Monday, April 3 | Observation |
| Tuesday, April 4 | Observation – Full period test on logs and exponents |
| Wednesday, April 5 | Observation |
| Thursday, April 6 | Observation |
| Friday, April 7 | No Observation – I was unable to attend |
| Monday, April 10 | Observation |
| Tuesday, April 11 | Observation |
| Wednesday, April 12 | No Observation – half day of school; class didn't meet |
| Thursday, April 13 | Observation |
| Friday, April 14 | Observation |
| Monday, April 17- Monday, April 24 | No Observations - spring break |
| Tuesday, April 25 | Observation |
| Wednesday, April 26 | Observation |
| Thursday, April 27 | No Observation – “Take your kids to Work” day; classes “canceled” |
| Friday, April 28 | No Observation – Field Trip; classes canceled |
| Monday, May 1 | Observation |
| Tuesday, May 2 | Observation |

During the time period of my observations, this class completed a unit on logarithms and exponents, spending one day on review and the following day taking a unit test. The test was very much like the review sheet they were given, addressing procedures and rules regarding manipulating logarithmic and exponential expressions

and equations. There were no application problems nor connections to graphs of such functions. The class then began a unit on rational expressions and equations. I observed the class for ten days of this unit, during which time the topics were simplifying, multiplying, and dividing rational expressions; finding a common denominator for rational expressions; solving rational equations; and finally adding and subtracting rational expressions. Additionally, during my observation period, Ms. Miller began giving out review packets on Mondays that were due on Fridays. Occasionally during class she would take questions on these. Following a brief description of how the time in this GT class was segmented, the discussion below addresses how Ms. Miller introduced and developed the unit on rational expressions.

As in the Algebra I class, Ms. Miller began each day with two drill exercises on the overhead projector. Most of the time, the drill problems reviewed past topics from the Algebra II curriculum. On about five occasions one of the two drill problems was on current material, presumably used to check that students were caught up with the material. All of the drill exercises I observed were straightforward symbolic algebra problems that required students to use a particular solution method. Students worked on drill problems individually, for about six to eight minutes. Then student volunteers would put their solutions on the board. Ms. Miller usually went through one or both solutions, talking out loud, and encouraging questions. On a few occasions, Ms. Miller asked students to question each other, rather than her, but this kind of interaction was not practiced consistently. This discussion part of the drill usually took from 3 to 10 minutes.

On several of the observation days, following the drill exercises, Ms. Miller took questions on the review packets she had distributed. These segments lasted anywhere from two to nine minutes. Before beginning the current day's lecture, she also would discuss the previous night's homework. She usually solicited student volunteers to put homework solutions on the board. On several occasions, she did most of the homework problems on the board herself. This was because the class informed her that they were unable to do them on their own. As with the drill exercises, she usually talked through the solutions that were on the board, soliciting questions from the students. The remaining class time was spent either with Ms. Miller doing example problems to teach a new concept or with students doing practice problems. Ms. Miller did not lecture per se; she seemed to teach simply by doing repeated examples of a technique, presuming that students would catch on to a pattern in her work.

On the third day of my observations, Ms. Miller began the new unit on rational expressions. On that first day of the unit, Ms. Miller gave a drill problem that required students to divide two rational expressions that contained only monomials. She said to the class while they worked on it, "I'm using number one to lead us into today." When she was ready to introduce the topic formally, she asked the students to look at the side board on which was written the "indicator/objective" from the curriculum. She asked the class if there were any words they did not understand. When someone responded that they did not understand the word "rational," she asked, "Does everyone know what 'expression' is?" She then told them to "go find out what a rational expression is." The students consulted their books. Then a student asked, "Does this 'rational' have anything to do with 'rational exponents'?" She responded hesitantly, "They're

similar...” and said nothing else about it. She went on to write the book’s definition of rational expressions on the overhead.

Next she wrote the following problem on the board that was of the same type as the drill problem: $\frac{8x^2}{26xy} \div \frac{12x^2}{14y^4}$. She gave the class two minutes to try this on their own and then began discussing it. She discussed multiplying by the reciprocal to “make it a multiplication problem because there’s a rule for this and there are rules in math which we must follow.” She also spoke about taking out common factors, but repeatedly referred to dividing as “canceling.” She pointed out to the students, “This problem is old news. We are using prior knowledge.”

She then asked, “What happens if I do this?” and gave them the problem, $\frac{x^2 + 6x + 9}{x^2 - 9}$, saying that they want to simplify the expression. She asked questions, but answered most of them herself to explain that they need to look for common factors. She said, “Factor it. Break it down into its simpler parts.” What ensued was confusion about what can and cannot be “canceled.” Students did not see the difference between factors and terms, and Ms. Miller did not demonstrate the relationship between rational expressions and simple fractions. She also failed to use mathematical language that might have helped the students. Rather than using the terminology of “factor” and “term,” she said things such as, “When we’re dealing with just single numbers or single variables you can cancel them...[we] can’t cancel there because it’s a binomial...we can only do it if it’s multiplication...you can’t break those two apart.” She put two more multiplication problems on the overhead. Students attempted to solve these examples and then watched her complete the solutions. Although she did encourage questions, her

answers did not seem to help students understand the mathematics. Many students could not do the examples on their own at all, and by the end of the class most students were still trying to divide out terms.

The next day, only three students were able to put homework problems on the board. After Ms. Miller put the rest of the homework problems on the board and tried to explain them, she spent the rest of the period giving the students problems to try as she went around and helped them individually. Based on conversations I overheard between her and individual students as she monitored their seatwork, it seemed that students were still having problems distinguishing terms and factors and therefore were incorrectly dividing out terms. The following day I could not attend, but when I asked her what they covered, she said they did more “simplifying and multiplying because they were having so much trouble -- they hadn’t even brought in division” yet. The next day, they spent half of the lecture time doing more multiplication problems before she introduced division problems. After explaining briefly that they would do a division problem by “turning it into a multiplication problem,” students were given two more such practice problems before class was over.

After spending four days manipulating rational expressions by factoring, reducing, multiplying, and dividing, Ms. Miller moved on to the next topic, which she wrote on the overhead as “Solving Rational Equations.” However, as will be illustrated in the description below, she was actually trying to teach students the concept of common multiples for polynomials.

In order to solve rational equations, it is useful to multiply both sides of the equation by a common denominator. Ms. Miller wanted students to learn how to

identify a common denominator as a first step. However, she never explained the utility and justification for performing that manipulation in solving rational equations, nor did she present a comprehensible way to construct a common denominator.

After writing “Solving Rational Equations” on the overhead, Ms. Miller wrote down her first example: $\frac{12}{8} + \frac{2}{3} = x$. She asked the students what she should do. She appeared to dismiss the student who kept suggesting that they reduce $\frac{12}{8}$, perhaps waiting for a suggestion about finding the common denominator. Finally, someone did suggest finding a common denominator, and she then wrote on the overhead: $\frac{36+16}{24} = x$. She then completed the problem, and simplified. A student seemed confused about how they had found equivalent fractions and suggested that they multiply $\frac{2}{3}$ by $\frac{1}{8}$ (instead of $\frac{8}{8}$). Once Ms. Miller realized what he was saying, she replied, “You need to multiply by a whole value of 1.” Several students also tried to help by saying, “What you do to the bottom, you need to do to the top.” Other students also seemed confused about equivalent fractions, and at least one student seemed to think that they multiplied the two fractions by different numbers, when in fact they multiplied by $8/8$ and $3/3$. Ms. Miller confused the matter by saying that she “multiplied one [term] by eight and one [term] by three.” (She meant $8/8$ and $3/3$.) She moved on to the next example even though many of the students did not seem to understand important concepts about fractions that would be needed to understand rational expressions. (Ms. Miller field notes / post-observation reflections, April 11, 2000)

Ms. Miller might have begun with simple fractions to tie some of the embedded concepts to problems involving algebraic fractions, yet she failed to address the misconceptions that her first example brought out. She also never addressed why 24 was the common denominator for 3 and 12.

Ms. Miller put a second example on the overhead that was very similar to the first, except for having three (rather than two) simple fractions on the left side of the equation. Without explaining how she arrived at the common denominator or how she

obtained equivalent fractions, she performed the operations to “solve” for x (which was already isolated on the right side of the equation). She then said, “The key here is to find the common denominator. Let’s talk about finding the common denominator.” They never actually talked about it at all; rather she continued on with more examples that in fact were rather complex.

She wrote a new heading on the overhead, “Common denominator,” and wrote the first example: x^2 and $x^3 - 1$.

She asked, “If these were my two denominators, what would be my common denominator?” There were no replies. She asked, “Do they share anything in common?” There were still no replies, so she continued, “Since they don’t share anything, you need all the pieces,” and she wrote on the overhead: “ $x^2(x^3 - 1)$.” (Ms. Miller field notes / post-observation reflections, April 11, 2000)

Students seemed to be extremely confused. It might have been due to their never having gotten a clear understanding of factors and terms over the past several days when they had been studying simplifying expressions. Additionally, her language, e.g. “anything in common” and “pieces” might have also made it difficult to understand the algebra. Ms. Miller then wrote “ $\left(\frac{4}{x^2} + \frac{3}{x^3 - 1}\right)x^2(x^3 - 1)$ ” directly underneath the common multiple $x^2(x^3 - 1)$. It appeared that Ms. Miller wanted the students to see that “things” would “cancel” when she multiplied that particular expression by the common denominator, but she did not show any equation. Students had no reason to be multiplying the large expression by the common denominator.

Ms. Miller then put a second example on the overhead in which she listed two polynomials, but asked for the common “denominator”, expecting students to factor

first. Once the class obtained the factors, students still had difficulty figuring out what the common denominator was. Ms. Miller did not probe the two incorrect suggestions that were offered. She tried to get the students to identify the common denominator by saying, “When we multiply the fractions by the common denominator, we want nothing left...in the denominator.” However, Ms. Miller had not made it clear that the polynomials she had written were denominators in the first place, nor had she explained what that would have to do with *multiplying* by a common denominator.

Since the class had been essentially non-responsive by this point, Ms. Miller wrote the answer to that problem, and went on to a third example. This example was a list of three monomials and a binomial. Ms. Miller said, “You gotta carry a lot of stuff. Do these share anything?.. Instead of taking the smallest, now I need the largest.” She talked through the problem quickly and wrote the common denominator on the overhead. The class period was over. This class would not be meeting the next day, since it was a half-day. Their homework was to do six problems in which they would need to identify a common denominator.

Overall, in the lesson described above, Ms. Miller did not tie the algebraic examples into her opening problems with simple fractions. She did not discuss with the class the meaning of a common denominator or common multiple and relate it to factors. The examples she used were rather complex, considering that her intended method of teaching this topic had been through illustrative examples. During the next class meeting, I heard students saying that the homework made no sense to them. One student said, “I got it, but the rules don’t make sense -- it’s like they just made it up.”

When Ms. Miller introduced the very next lesson she said, "... we moved into adding and subtracting rational expressions. We're focusing on finding the common denominator. After we go over homework, we'll try to move into solving those expressions." After a homework discussion made it clear that most students were still very confused, she said, "Let's move onto problems and we'll use our homework to practice." What followed were more examples in which Ms. Miller solved rational equations. The next day Ms. Miller spent a lot of time doing the homework problems at the board herself, since students were having so much trouble. The remainder of that day was spent with students working individually on more examples of the same topic, solving rational equations, as Ms. Miller went around to help. This was the last day before Spring Break, when students had six school days off.

When students returned from Spring Break, the next two days were spent reviewing all of the topics they had covered in rational expressions, by way of doing more practice problems. Most of the time students worked individually on these problems, as Ms. Miller walked around giving help, but she also did some of the problems on the overhead projector, soliciting help from students. On the second day back from break, after spending nearly 30 minutes of the period on simplifying, multiplying and dividing polynomials, she said, "We need to get caught back up on addition and subtraction." However, they actually had been solving rational equations. They then spent 15 minutes reviewing finding the common multiple (although she called it a common denominator) for two or three polynomials and then solving rational equations. The next two days of class were canceled for different reasons.

On the next day that the class met, once again the bulk of class time was spent with students individually practicing solving rational equations. Ms. Miller intermittently put solutions on the overhead projector. No new material was presented. The next day, which was the fourth class meeting after Spring Break, Ms. Miller moved to the next topic in the chapter, which was adding and subtracting rational expressions. She spent 27 minutes doing examples on the overhead as a means of presenting the new concept. Throughout the “lecture,” the explanations Ms. Miller provided for the procedures were characterized by imprecise language and errors. For instance, Ms. Miller asked the class, “What kind of steps can we write down for unlike denominators?” She wrote on the overhead, “1) Find the common denominator; 2) Multiply the ...” Ms. Miller got stuck here, saying, “I can’t think of the word I want here. What are we doing?” A student suggested something about multiplying the numerator and denominator. She replied, “We don’t want to multiply the denominator because we already said, ‘Find the common denominator’ ... This is really bothering me...well, I won’t fill it in -- we’ll just say, ‘Multiply the numerator.’” Ms. Miller was struggling with equivalent fractions and unit fractions, or fractional equivalents of one, neither of which she mentioned during this lesson. That lesson was on my final observation for this class. However, I did observe the Algebra I class on the following day, so I asked Ms. Miller what they had done in the GT class that day. She responded, “Their homework from last night was a nightmare -- they were so frustrated. They could find the common denominator, but didn’t know what to multiply by.” She said they spent the whole period just working examples.

The chapter that Ms. Miller had covered was actually termed “Rational Functions” in their textbook. However, Ms. Miller had never discussed rational functions, nor tied any of what they were doing to graphs of rational functions. The material she taught was purely mechanical. I am not sure why she chose to emphasize just those topics that the district was actually trying to de-emphasize, as opposed to a more integrated graphical and algebraic approach, as their text book presented.

Ms. Miller’s Instructional Practices

Tasks

In the Algebra I class, I observed Ms. Miller using many tasks that at face value had the potential to engage students in meaningful mathematics. For example, the carnival activity and numerous simulations provided real-life contexts, as well as posed compelling questions that required thoughtful engagement and investigations to answer. Additionally, I observed Ms. Miller conducting a whole-class lesson using an application problem that dealt with some geometry and a quadratic equation.

However, as discussed earlier, Ms. Miller did not implement the activities in ways that might have fostered the students’ careful examination of the mathematics at hand. Although she solicited ideas from the students during the whole-class segments of the lessons, her responses and questioning fell short of developing exchanges with students that took their original ideas into account. In fact, she often failed to recognize correct ideas or responded in a dismissive manner seemingly because she had particular ideas in mind that she appeared determined to impart.

For instance, in one simulation exercise students had to determine if a neutron would hit the end of a finite grid (“the wall”) within five seconds when, at the end of each second, it moved in one of four directions on the grid. While students were carrying out the simulation, I heard her say to a group in an annoyed tone, “You need to spin it five times [per trial]!” One of the students said, “But if it already hit the wall...” Ms. Miller abruptly said, “Regardless, you need to spin 5 times so we can see what else you did. I can see your logic, but spin it 5 times.” In fact, it could have made for interesting data to record the information the student alluded to -- how long it took for the neutron to hit the wall.

On other occasions, Ms. Miller illustrated limited conceptual understanding, which also affected her ability to implement a task or activity to its full potential. Near the beginning of each simulation activity, one of Ms. Miller’s goals appeared to be getting the students to identify the tool they should use for the simulation, e.g. coin, die, or spinner. She began a discussion by asking, “What two objects have we used to represent probability?” She waited for the desired answers of coin and die. Then she asked, “If I want to play evens/odds, what should I choose?” Someone said dice, perhaps because there are even and odd numbers on the faces and because in fact when they did the carnival activity two days prior, the “Evens or Odds Booth” required students to roll a die; they would score points if they rolled 2, 4, or 6. She ignored this response entirely, and said, “Coin, because there are 2 outcomes on a coin.” She then asked the class, “What can I use if I want to simulate something with three outcomes?” After many hints, someone finally responded in a reticent tone, “Spinner?” It seemed that this was the answer Ms. Miller had wanted. In fact, a die could have been used if

one assigned two numbers to each of the three outcomes. It appeared as though Ms. Miller had certain ideas in mind that she wanted to convey to her students without being diverted. It could also be that she confused the issues of *equal* outcomes with *number* of outcomes.

In the GT class, I did not observe Ms. Miller using any applied tasks. During the entire observational period, which spanned five weeks, Ms. Miller's lessons comprised only abstract, symbolic (as opposed to graphical or contextual) algebra problems, from drill exercises to class examples and practice problems, to homework problems. The only tasks students had engaged in were watching Ms. Miller manipulate expressions and then practicing their own manipulation skills in a "do-as-I-do" lesson format.

Learning Environment

There were certain surface elements, such as the non-traditional arrangement of desks in her classroom and the ease with which many students interacted with her, that might have helped Ms. Miller establish a positive learning environment. Yet, the learning environment was far from the ideal put forth by NCTM. That is, as a result of the content presented to the students and the pedagogical practices employed by Ms. Miller in both classes, mathematics was conveyed as a set of rules to be learned, rather than as a sense-making pursuit.

Ms. Miller had the desks in her classroom arranged in a manner that might have allowed students to interact easily with one another and her. Yet, in both classes, interaction between students was limited. The Algebra I class was large enough that indeed, students had to sit adjacent to other students in most cases. During simulation

exercises, students worked in groups (usually where they were already seated), and still had a view of the front of the room for whole-class discussions. However, students were observed working together only to carry out physical tasks such as rolling dice or spinning spinners. They were never asked to work with partners or groups to discuss, resolve, or reconcile mathematical problems of any kind. The GT class had so few students to begin with, that no one sat directly next to anyone else, whether by each student's own choice or the teacher's. Regardless, this arrangement surely made interaction more difficult. Indeed, the only interaction that was observed was informal, in that a student might privately ask another student for help, usually as class was getting started, or as students were waiting for a new lesson segment to begin.

Clearly in both classes, students seemed to feel comfortable speaking out and asking questions. When Ms. Miller asked students if they had questions on their review packets, there was always a response. Students often volunteered to put a drill solution on the board, even when they had only partially completed the problem and needed some guidance by Ms. Miller to complete it. In the GT class, several times students readily admitted that they were unable to do their homework and asked Ms. Miller to do the problems on the board and give them more practice. However, the content that Ms. Miller taught and the manner in which it was presented, was extremely procedure-laden. As a result, students generally interacted within this agenda and thus, focused their own efforts and inquiries on procedures and rules.

Indeed, where a positive learning environment should have stressed that there are reasons behind rules, it often appeared that Ms. Miller wanted to convey that rules underlie the reasons. For instance, when she went over a problem that involved division

of rational expressions, she asked the class, “Why did he make it a multiplication problem?” She answered the question herself, saying, “There’s a rule for this and there are rules in math which we must follow...” On another occasion, she explicitly deferred to “the rules” aloud perhaps out of frustration. She was trying to explain why one could not factor $5x$ out of $5x^2 - 20$. She said, “You can’t take out an x because 20 doesn’t have an x ...because it has to be common to both...because those are the rules that we learned!”

Discourse

As alluded to in the section on learning environment, Ms. Miller set up a pattern of discourse in her classes that might have been more productive in terms of helping students develop reasoning and conceptual understanding of mathematics, had the focus of that discourse been different. That is, most of the questions that Ms. Miller asked and elicited from her students were procedural in nature, and responses on her and her students’ parts usually fell short of making mathematical relationships explicit.

Ms. Miller’s lecture-style incorporated questioning her students, but those questions rarely touched on concepts, even if the concepts underlay the procedure she was teaching. The details of lessons outlined in previous sections indicated the limited nature of her questions and responses. As another example, consider the instance in which one of the GT students asked a pointed question about the algebra in solving a rational equation. In trying to clear up confusion that many students in the class were having about multiplying an equation through by a common denominator, the student asked, “Even though the “5” and the “ $x - 4$ ” are on different sides of the equal sign, you

still multiply both [sides] by both [factors]?” Rather than seizing an opportunity and expounding on the reasoning behind this manipulation, Ms. Miller answered only, “Yes.”

There were exceptions to her more usual line of procedural questioning. For example, in explaining to a student the equations for the graphs of horizontal or vertical lines, she asked, “If I draw a line like this [horizontal], what is staying the same? What does every point on this line have in common?” After doing a review problem in Algebra I that consisted of a linear graph in a specific context, she asked, “What was the y-intercept?” When students responded with a number, she said, “Be more specific -- in terms of this problem, what does this mean?” Although these are examples of more conceptual questioning patterns, unfortunately these kinds of questions were not consistently asked across content matter. One conjecture could be that Ms. Miller was better able to ask conceptual questions with material with which she, herself, was much more comfortable.

In both classes, Ms. Miller had students put drill and homework solutions on the board. Although there were times when she purposefully directed a student’s question at another student, or had a student explain their work on the board, these practices were not done consistently. More often than not, she, herself, would talk through a student’s solution that was put on the board. Furthermore, the students that put work on the board had volunteered to do so. It became apparent that in the GT class, there were particular students who almost always put the problems on the board. These same students asked many questions. On the other hand, there were three other students that were consistently silent; they never asked questions, nor volunteered to answer any. On at

least one occasion, I heard three students having a productive-sounding discussion about how to add rational expressions by finding equivalent fractions (although, they were *not* heard using that particular vocabulary). This had been a trouble spot for the class. Ms. Miller seemed to notice the conversation going on, did not comment on it, or ask them to share their ideas with the rest of the class. This kind of failure-to-facilitate points to Ms. Miller's inability, or perhaps unwillingness, to cultivate the kind of discourse that could help students make sense out of mathematics.

Other Themes

The data collected during my observations of Ms. Miller illustrated that she had gaps in mathematics content knowledge and used imprecise language when teaching algebraic concepts. Additionally, there were several occasions on which she failed to explain errors she had made in a previous lesson.

Content Knowledge

Ms. Miller exhibited poor mathematics content knowledge in areas of elementary and intermediate algebra. In Algebra I, she seemed unaware of the significance of the zero-product rule and in GT, she frequently used imprecise, or incorrect terms, when explaining algebraic procedures. Ms. Miller chose to focus narrowly on particular skills. She did not tie any of the observed algebra topics -- quadratic equations, or rational expressions and equations -- to graphical representations.

Throughout the probability unit in Algebra I, she seemed unaware of the fundamental concepts that some of the activities had embedded in them, such as the law of large numbers or equal versus unequal outcomes. Perhaps because she was unaware of these concepts herself, she was unable to engage students in productive, meaningful discussions of them. Generally, there was very little closure on the simulation activities.

Many of the characteristics of Ms. Miller's instruction, such as the absence of a conceptual focus, insufficient explanations, and nearly sole reliance on teaching through examples as a methodology, might reasonably be traced to a lack of adequate mathematics content knowledge. The data seem also to reveal that Ms. Miller might hold certain beliefs about the role of rules in mathematics that would have the same narrowing effect on her teaching.

Admitting errors

On three occasions, Ms. Miller learned of errors that she had made during her Algebra I class. Each of those episodes was described previously. As mentioned, Ms. Miller did not attempt to explain her mistakes to the class, nor did she acknowledge her error when she became aware of it *during* the lesson itself. In each case, students were left with incorrect information and possibly the seeds of more serious misconceptions. Furthermore, if students analyzed the information presented so that they recognized the lack of mathematical consistency, it might have reinforced the notion that mathematics need not make sense.

Comparison of Ms. Miller's Algebra I and GT Classes

As in the other two teachers' classes, Ms. Miller's GT class was very small, having only 11 students as compared to the Algebra class, which had 25. The small number in the GT class allowed Ms. Miller to provide more individual attention to students in that class. However, this difference in interaction with individuals could be explained by a separate finding. That is, students were engaged in very different kinds of work in the two classes. In GT, Ms. Miller gave students quite a bit of time to work on practice problems individually. During this time, she went around to help them. In Algebra I, students either listened in a whole-group setting as Ms. Miller introduced a simulation experiment, or they worked in groups to carry out the simulation. However, Ms. Miller did give students some individual attention during the time she gave them to work on their review packets. Interestingly, she only gave the Algebra I class time to work on review packets *during* class. In the GT class, she only addressed questions they had from working on the packets on their own time. This difference could be due to a lack of confidence that her Algebra I students would complete their packets if not given time in class.

One element that became clear from the data was that in the Algebra I class, Ms. Miller often got visibly angry at the class, and spent several minutes on at least four different days reprimanding the whole class for their lack of effort, or the poor quality of work, and admonishing them about the consequences. This did not occur in the GT class. On the contrary, she seemed genuinely sympathetic when they were having trouble grasping certain skills, and was particularly sensitive to one student who had obvious insecurities about his abilities. On several occasions, she would give him praise

or a “pep talk,” and encourage him and any other students to continue to come for extra help during lunch. This difference in treatment of students could be due to the fact that indeed the GT students took more responsibility for their learning, or to the perception that they were better students by virtue of their being in GT.

PA Alignment with Instruction and Opportunity to Learn

The Algebra I class did engage in activities that were aligned with the PA. That is, the probability experiments required students to analyze a problem, set up a reasonable simulation to answer the question, and then interpret results. The summary questions were open-ended and required students to write out an appropriate response based on their work. However, as described previously, whereas the activities possessed many similar elements to the state PA tasks, they were not implemented in a way that would clarify meaning or conceptual relationships. The teacher, rather than the students, established a course of action, and students were not asked to think through the key mathematical concepts of the activities. However, throughout the activities, students were reminded of certain elements of an appropriate answer, as deemed by the PA, such as well-labeled graphs, and complete and sensible sentences.

The lessons I observed in the GT class did not seem to possess any characteristics that would be commensurate with PA expectations. There was almost no concept development for, or reasoning through, the algebraic content they were covering.

Discussion of Research Questions

Based on the analysis of the data from three observed eighth-grade mathematics instructors in Arbor District, I will discuss below all three main research questions I sought to answer through this observational study.

Research question (a): *What are the characteristics of instruction in the classrooms of eighth-grade mathematics teachers identified as “exemplary” traditional and reform teachers?*

Ms. Drake and Ms. Miller were both identified by their principals as exemplary. Ms. Drake was acknowledged to have a more traditional style, whereas Ms. Miller was described with attributes that were more in line with current reform efforts, such as giving activities that require writing and explaining, or are extensions of previously considered work. The specialist in the District Office of Mathematics identified Ms. Henderson as a teacher who had incorporated many reform approaches into her teaching practice, while maintaining a traditional style of teaching.

In all three classes, there were no notable classroom management problems. Both Ms. Drake and Ms. Henderson had pre-emptive classroom procedures that eliminated time-wasting activities such as finding and sharpening pencils or turning in work. Additionally, all three teachers had a well-established routine, at least for the start of every class, whereby a drill exercise was projected on an overhead before all of their students had even arrived.

All three teachers used activities with at least one of their classes that had the potential to encourage students to connect mathematical relationships, and to ask them

to reason through non-routine problems. Many of these activities resembled the kinds of tasks commonly found on the State performance assessment.

Both Ms. Henderson's and Ms. Miller's students were used to being asked questions by their teachers and asking questions of their teachers. Both classes had a considerable amount of interaction. Ms. Henderson's classes were dominated by constant interaction, and she often directed students' questions back to other students. All three teachers had students in at least one of their classes work in groups on some occasions, although only in Ms. Henderson's classes was this a noticeably formalized practice.

However, an important finding based on the observations described in this study is that principals may not be the best-suited individuals to assess the efficacy of mathematics teachers. That is, both Ms. Drake and Ms. Miller had serious deficiencies in their mathematical content and pedagogical knowledge that their respective principals might not have recognized. In both cases, these deficiencies translated into limited mathematics lessons and a less than ideal learning environment that superseded the positive characteristics described above.

Although in Ms. Miller's classroom there was a considerable amount of interaction, the discourse was too often focused on procedures and rules, with little if any emphasis on conceptual relationships in mathematics. In Ms. Drake's class, there was very little interaction altogether, and more notably, almost none in terms of student-initiated questions. As suggested by Artzt and Armour-Thomas (1999), a teacher-directed style of teaching, including restricting the kind of discourse that occurs in a classroom, "can serve as a mask for teachers who do not possess full knowledge of the

content, student, or pedagogy”(p. 229). Thus, although Ms. Miller and Ms. Drake possessed some qualities that had potential to render exemplary mathematics teaching, their deficiencies in mathematical content knowledge prevented them from realizing that potential.

Ms. Henderson, however, exhibited many characteristics of exemplary teaching, most of which were described above. Ms. Henderson used hands-on tasks with her students and used questioning techniques to highlight the significant mathematics content embedded in those tasks. She utilized graphing calculators, but as a tool of exploration, rather than an end in itself. In teaching algebraic principles, she incorporated graphical approaches that helped students make connections among the various mathematical relationships and representations. Perhaps the most seemingly effective characteristic of Ms. Henderson’s teaching, and the one that stands in sharpest contrast to the other teachers observed, was the amount and intellectual level of discourse that permeated all of her lessons in both classes. Ms. Henderson’s questioning, as well as the kinds of questions observed coming from students, were both procedural and conceptual in substance. For example, students in the GT class commonly asked very exacting questions aimed at understanding the concepts behind algebraic manipulations, while in the Pre-Algebra class Ms. Henderson persistently asked the students to recognize and explain geometric relationships in a variety of situations.

The district mathematics specialist used both words -- “traditional” and “reform” to describe Ms. Henderson. Notwithstanding the pedagogical “baggage” that these

labels inevitably carry, Ms. Henderson clearly exemplified effective instructional practices.

Research question (b): Are there any comparisons or contrasts in the high-track and low-track classrooms with respect to teaching for conceptual understanding, particularly for two levels with the same teacher?

All three teachers were observed teaching Algebra II (GT), the highest-level eighth-grade course in the district. Both Ms. Drake and Ms. Henderson were also observed teaching the lowest level eighth-grade courses for their schools. Ms. Miller was observed teaching Algebra I, which was the middle-level course for eighth grade at her school. In all three cases, the GT classes were considerably smaller than the other classes in which each teacher was observed. The GT classes had 14, 16, and 11 students, compared to 23, 27 and 25 students, respectively. Class size might determine the amount of individual attention any student receives from the teacher, as well as the amount that any individual might be able to orally contribute to the class. However, if class size influences the nature of the activities and interaction that occurs in a classroom, it was not found to consistently favor the smaller GT classes across the three teachers.

In Ms. Drake's class the kind of interaction observed was so limited in both classes, that no noticeable differences were detected. However, tasks presented to the GT students were unquestionably more mathematically rich than the lessons presented to the students in the Fundamentals class. The GT students were given activities that could have given them the opportunity to employ mathematical skills such as reasoning,

conjecturing, finding relationships, and proving, had the lessons been developed differently. On the contrary, the students in the Fundamentals class were relegated to abstract lessons and routine tasks that intrinsically did not even possess the potential to help students develop conceptual understanding in mathematics.

Unlike the situation in Ms. Drake's classes, it was Ms. Miller's GT class that received much more routine and abstract mathematics lessons than the lower-level class. The only tasks that GT students were observed doing were watching Ms. Miller complete routine algebraic manipulations and then trying similar problems on their own. The Algebra I class, on the other hand, completed many activities that were rich with the potential to allow students to examine meaningful mathematics. However, Ms. Miller's presentation of these tasks failed to highlight the substantive concepts and connections that were embedded in them. In Ms. Miller's Algebra I class, as in Ms. Drake's GT class, the teacher, rather than the activity, was the obstacle to helping students learn meaningful and connected mathematics.

Both levels of Ms. Henderson's classes abounded with interaction that emphasized mathematical concepts as well as procedures. Ms. Henderson made it a point to have all students participate, so that students in the GT class each might have been able to participate more, simply because their numbers were fewer. Aside from the mathematical content of the courses being different, Ms. Henderson did not appear to have any obviously different expectations for her students with regard to the kind of mathematics they were capable of learning. She gave both classes activities that contained significant mathematical content.

Ms. Miller did give the students in the GT class much more individual attention than she did the students in the Algebra I class, due to the nature of the lessons she presented to each class. However, because the focus of this attention was on manipulation and procedures, it is unlikely that the GT students gained more conceptual understanding from the interaction. It was also noteworthy that Ms. Miller reprimanded her Algebra I class on several occasions, but did not show this kind of frustration with the GT students, even though she spent many days repeating the same information. This could be a manifestation of different expectations she had for students in the two classes. It appeared as though she entrusted the GT students with more responsibility for their own learning. On the contrary, she might have believed that the Algebra I students needed to be threatened into studying more.

Research question (c): Were the instructional activities and discourse aligned with the expectations evident on the state performance assessment? That is, did students in the classrooms observed appear to have the opportunity to learn the skills assessed by the mathematics component of the PA?

In Ms. Henderson's classes the interaction that was observed and the tasks in which students in both classes were engaged appeared to be well aligned with the State PA. That is, students were expected to reason, explain, justify, and conjecture about the mathematics they were learning. Furthermore, students frequently worked in pairs or groups with clear expectations about how to use time productively. In both classes, Ms. Henderson gave activities with extensive directions and formally discussed with students strategies for comprehending and carrying out instructions.

Ms. Miller's two classes differed from each other, as well as from Ms. Henderson's with respect to the amount of exposure each had to the kind of learning that might help their performance on the state PA. The mathematical content the GT students were studying and the way in which the mathematics was presented during my observations appeared to share nothing in common with any mathematics that might be assessed on the PA. On the contrary, the Algebra I students engaged in many exercises that had typical elements found on the PA: context-bound mathematics, proposing a plan of action, gathering data, analyzing data, and explaining results. However, as discussed in previous sections, the limited way in which the activities were carried out and discussed calls into question what students might have gained from doing these activities. However, students in both of Ms. Miller's classes were accustomed to asking and answering questions, even if the emphasis of the exchanges was procedural in nature. Despite the lack of conceptual discussions, students still had opportunities to verbalize mathematics, a skill assessed on the State PA.

In Ms. Drake's classes, only the GT students were exposed to activities that shared any common elements with those on the PA. However, as in Ms. Miller's Algebra I class, the lessons were implemented in a very narrow way, limiting the kind and amount of mathematics that students might learn from them. At least as important as the tasks in which students were engaged was the kind of interaction that took place in the classroom. In both of Ms. Drake's classes, the discourse was extremely limited. Students were almost never asked to explain mathematics, or think about it deeply. For these reasons, Ms. Drake's instruction did not appear to be congruous with the state PA.

Survey Scores and Observed Instruction

A comparison of the survey scores that the observed teachers received on reform instruction, along with a detailed description of their teaching styles provided in this chapter shed some light on the validity of the survey score itself. Each teacher who was observed answered the 24 items on instructional strategies three separate times: twice for their observed and surveyed class during the regular survey mailings in November and March, and once for the observed, but non-surveyed, class. The latter was filled out during the course of the observational period at each teacher's convenience. Table 25 gives the scores that the teachers received on those three occasions. The November and March scores for the surveyed class were averaged; that number is listed in the third column in the table.

The numbers in Table 25 illustrate that the reform score as composed from the questionnaire items on instructional strategies might have captured teaching styles on a spectrum of traditional to reform, at least relative to other teachers. That is, for the most part, the numbers in the table make sense relative to one another. For instance, Ms. Drake's Fundamentals class, which was described as having no elements in common with typical reform recommendations had the lowest score of all, .25. On the other hand, both of Ms. Henderson's classes had very high scores, with Algebra II having the highest score of all, .67. Recall that Ms. Henderson's classes were characterized by the use of instructional strategies commensurate with reform ideals: context-bound problems, calculators for exploration, and student-derived, well-reasoned written and oral justifications. Both of Ms. Miller's classes had reform scores that were on the lower end of the scale, although not as low as Ms. Drake's Fundamentals class. The scores

Table 25
Reform Scores for Observed Teachers

| Teacher Course | Reform Score | Mean Score for Surveyed Class |
|------------------------------|--------------|-------------------------------|
| Ms. Drake | | |
| Fundamentals ^a | .25 | |
| GT (Algebra II) - November | .65 | .55 |
| GT (Algebra II) - March | .45 | |
| Ms. Henderson | | |
| GT (Algebra II) ^a | .67 | |
| Pre-Algebra – November | .61 | .575 |
| Pre-Algebra – March | .54 | |
| Ms. Miller | | |
| GT (Algebra II) ^a | .36 | |
| Algebra I – November | .33 | .375 |
| Algebra I – March | .42 | |

Note. Scores represent percent of class time using 14 reform strategies out of 24 total strategies listed.

^aClass was observed, but surveyed only on instructional strategies.

that were most problematic were those of Ms. Drake’s Algebra II class. First, those two scores were much further apart than the other two teachers’ two scores from the same class were. That is, Ms. Drake’s GT class had two scores with a 20% difference, whereas the other two teacher’s scores had differences of only 7% and 9%. The second point is that Ms. Drake’s GT class received relatively high reform scores with an average of .55. One explanation is that indeed, Ms. Drake did use many reform

strategies, if only nominally so, in this class. That is, during the course of my own observations, her GT class did in fact engage in several extended problems that were context-bound, integrated algebra with geometry, required the use of graphing calculators for finding patterns, and required constructed, written responses.

What these numbers reveal is the limitation of this study's survey to capture the spirit with which various instructional strategies were carried out. As discussed in this chapter, classroom discourse, in whatever form it takes, is a critical element of the mathematics learning environment. Discourse is the medium through which mathematical meaning can be made, or not, out of the mathematical tasks students work through in class. Ms. Drake's high reform scores for Algebra II likely reflect *what* she did, but were unable to take into account *how* she did it. As described in that section, the tasks she brought to that class were mathematically rich, but the potential they possessed was lost through the lack of concept development, cohesiveness, and discussion.

CHAPTER 6: CONCLUSIONS

This study sought to accomplish several tasks. The main goal of the quantitative study, which utilized a teacher survey as a principal data source, was to determine the kinds of instructional strategies eighth-grade mathematics teachers perceived they employed, and whether teachers of three different academic tracks, or levels of courses, used different kinds of instructional strategies in a predictable pattern. One of the key questions was whether and in what way student achievement levels on the State performance assessment were influenced by the amount of reform instruction that their eighth-grade teachers employed.

Furthermore, the qualitative component of this study sought to describe in detail the instructional styles of three of the surveyed eighth-grade teachers who were identified by a superior as being exemplary teachers. The observations of those teachers served several purposes. First, the descriptions provided accounts of actual instruction that took place in six classrooms in the study district, as opposed to a questionnaire that could only give an estimate of instructional strategies based on the responding teachers' perceptions. Second, the observations offered insight into the complex nature of the classroom, and the kinds of classroom and teacher characteristics that ultimately support or hinder teachers' efforts to develop their students' conceptual understanding in mathematics. Finally, the observations provided a glimpse into the validity of the survey score on the use of instructional strategies and the efficacy of using a questionnaire such as the one used in this study to assess instructional practice.

Below, I discuss the results of both the quantitative and qualitative components of this study in the context of mathematics education in general, from the standpoint of

a constructivist philosophy of learning. The extent to which the original research questions were answered, the implications of this study, and the limitations inherent in this study are the focus of the discussion.

Implications

Assessment of Instructional Practice

This study utilized two methods to assess teacher's instructional practice. A series of items on the teacher questionnaire provided a quantitative measure of 51 teachers' preferred instructional style on a traditional-to-reform scale. Observations of three teachers over a three-to-five week period offered detailed descriptions of instructional practices in six classrooms. As prefaced in the Methods section, each method of assessing instructional practice has pros and cons.

Observations, and subsequent recordings of one researcher, have the advantage that they do not depend on individual perceptions of the teachers involved. That is, the subjects, themselves, need not make any judgments on what they do, nor must they interpret what a particular description of an instructional strategy might mean. In this study, the same researcher observed six classrooms. Although not wholly objective, the lens through which I viewed what transpired in the classrooms was the same lens in all classes. Furthermore, that lens was described in the Observations section. I believe that the accounts of those classrooms provided in this study gave an accurate picture of what those teachers regularly did and did not do with their students. Although the qualitative data were not quantified in a manner that could be directly comparable to the survey score, readers familiar with issues in mathematics education could evaluate the

instruction of those teachers, individually, and in comparison to one another.

Unfortunately, the method of using an extended observational period requires a tremendous amount of time and is prohibitive when a large sample is necessary.

In order to obtain a reasonable sample of teachers, a questionnaire was necessary. Teachers were asked to circle a number from 1 to 5 that most closely corresponded to the frequency with which they used a particular instructional strategy. Teachers also needed to estimate the number of minutes on average in a class period during which their students would be engaged in that particular strategy. Having teachers estimate the number of minutes, rather than circling a choice of given approximations, was thought to be an improvement on a previously used questionnaire that had the same goal (Mayer, 1998). However, it is possible that estimating minutes was not as productive as was anticipated. Several subjects were dropped from the study because they did not answer those questions, and comments from other teachers led me to believe that estimating the number of minutes was too confusing or difficult for some of the teachers. Because of that, it is likely that there is measurement error in the final scores.

Furthermore, as alluded to above, a questionnaire inherently requires interpretation on the part of the respondent. Not only were teachers reporting their perceptions of what they did, but they also were interpreting the particular items. On the questionnaire, for example, one of the items asked teachers to report how often and long their students “work together in pairs or small groups on mathematical problems.” A teacher might interpret that to include having their students pair up to check homework answers. Yet another teacher might not consider that strategy to include answer-

checking, but instead interpret it more narrowly to include only times when students were working on a new problem that required idea-sharing and a student-derived course of action.

Nevertheless, the reform scores based on the survey were used to address the first research question: What factors related to teacher, classroom, or student characteristics were predictive of the amount of reform usage that teachers perceive they implement? Only five possible predictors were examined due to limitations imposed by the sample size of 51, and of those five, only years teaching grades 7-12 entered the regression equation. The unstandardized coefficient for that variable was small; each year of teaching experience in grades 7-12 yielded an expected increase in reform usage of just over a half of a percent of class time.

Another drawback of the survey as a way to assess classroom practices became very clear from one of the observed teachers' score in particular. Ms. Drake's two reform scores for her Algebra II class were relatively high, which was inconsistent with the classroom atmosphere and the actual content that I observed. As explained in chapter 5, although her methods could be nominally labeled as reform, she implemented those strategies perfunctorily, and thus ineffectively. While there were other facets of her instruction that possibly limited students' conceptual understanding of mathematics, such as teacher content knowledge, the overwhelming aspect of reform instruction missing from this class was productive discourse. The survey items were unable to adequately capture an essential ingredient in classroom instruction – meaningful interaction focused on mathematical concepts. That is, although there were three instructional strategy items from the questionnaire that most directly addressed

discourse, perhaps those alone could not account for the vast difference in quantity and quality of discourse evident in Ms. Drake's and Ms. Henderson's classes.

Those three questionnaire items that addressed discourse asked how often students (a) "orally explain how to solve a problem," (b) "make conjectures and discuss various methods during problem solving," and (c) "engage in student- or teacher-led whole group discussion." However, as was evident in Ms. Henderson's classes, student verbalizations of mathematics can and should permeate the entire instructional process. In Ms. Henderson's classes, except for time spent on individual tests, students were expected to orally explain, reason through, or justify every mathematical statement from opening problems to closing discussions. Furthermore, they were consistently encouraged to question the teacher and each other. This kind of discussion and interaction is considered an essential component in teaching mathematics for understanding (Brooks & Brooks, 1993; Driver, 1995), a notion based on the constructivist and conceptual change philosophies. That is, conceptual change ultimately depends on metacognitive activity, which is fostered through interaction and reflection (Hewson, 1996; Underhill, 1988).

Along those lines, it is important to keep in mind that constructivism, a theory about how knowledge is constructed, maintains that learners construct knowledge through their experiences, no matter what those experiences are (Cobb, 1994). That is, no matter how reformist or traditionalist a classroom setting or teacher is, students will go through processes to construct knowledge. The question is: What kind of knowledge is being constructed? Research needs to continue to focus on the learning environments, including instructional strategies, that best support the construction of mathematical

understanding. Additionally, it would be informative for the field of teacher education to better understand what kinds of pre- or in-service experiences are associated with various learning environments that teachers establish.

Teaching Credentials, Teacher Content Knowledge, and Instructional Practice

The teacher survey afforded the opportunity to examine differences between teacher and classroom characteristics in different eighth-grade course levels, which was the focus of the second research question. The one clear disparity that was evident was the varying levels of professional credentials held by teachers in the three tracks. Forty percent of the Pre-Algebra teachers, which included the Math 8 teachers, were lacking any substantial mathematics preparation, compared with only 12.5% of Algebra I teachers. Those teachers held either provisional, emergency, non-mathematics middle/secondary, elementary education, or special education certification. All of the surveyed Algebra II teachers held the highest level of certification, which was middle/secondary school certification in mathematics. A reasonable concern is whether teachers who have less knowledge of mathematics – that is, many of the Pre-Algebra teachers – are less able to impart meaningful mathematics to their students through effective pedagogical practices that might be tied to mathematics content in particular. Indeed, studies have found significant differences in teacher knowledge and teacher skills that translated into student achievement between certified and non-certified teachers (Hawk, Coble, & Swanson, 1985).

The Pre-Algebra students in this study did score significantly lower on the state performance assessment than students in either of the algebra courses. However, results

from this study did not conclusively indicate that Pre-Algebra teachers used different teaching strategies than teachers in the other two course levels. Teachers' instructional styles were measured on a continuum of reform approach. That is, each teacher received an average reform score that reflected the percent of classroom time spent on 14 reform strategies, as opposed to traditional strategies. For this measure, there were no significant differences found between the three course levels. Therefore, this study did not identify course level as an indicator of the kind of teaching that reportedly occurred in the classrooms. Furthermore, the teacher observations depicted two teachers who *did* hold certification in middle/secondary mathematics, and yet had serious deficiencies in either mathematics content knowledge or pedagogical content knowledge or both.

Without more in-depth teacher interviews, it is difficult to separate the two.

Nevertheless, a post hoc analysis showed a link between teacher credentials and student achievement, while controlling for course level. The results were in the expected direction, with higher levels of certification yielding higher achievement on the PA. That finding corroborates results from previous studies that found direct (negative) correlations between teacher competencies, as measured by standardized tests, such as certification tests for teachers and student failure rates on standardized high school examinations (Strauss & Sawyer, 1986).

Thus, although there was an inequitable distribution of teachers with appropriate mathematics credentials in the surveyed classrooms, it is not clear how that finding translated into mathematics instruction to which students were exposed. With the initiation of the No Child Left Behind legislation, school districts are currently clamoring to put "highly qualified" teachers into every classroom. If they do, the

finding from this study would be deemed an unfortunate artifact of the past. However, until that requirement is accomplished, and done so without the aid of loopholes, it would be a worthwhile research endeavor to investigate the impact of various teacher credentials on what teachers do in the mathematics classroom.

Furthermore, Pre-Algebra teachers reported the highest score for the match between the PA and their teaching style. The PA is a constructed-response performance assessment that uses real-life contexts as backdrops for mathematical problems. Its problems require students to analyze, interpret, and synthesize information, and then to communicate reasoning. Therefore, because of the nature of the PA, it might be reasonable to assume that a match between PA and teaching practice implies a reform approach to teaching. However, as already discussed, the Pre-Algebra teachers did not have a significantly different reform score than that of the other teachers. Additionally, a closer examination of the individual instructional strategies revealed that Pre-Algebra teachers reportedly employed a smaller *variety* of reform strategies among their most-used instructional strategies. The reported frequency and duration of reform strategies aside, another hint at what might have actually occurred in many Pre-Algebra classrooms might be gleaned from the fact that Pre-Algebra teachers reported a significantly greater influence of the FMT on their instruction as compared to teachers in the other two courses. The FMT is a multiple-choice, basic skills test emphasizing computation with fractions, decimals and percents. It would be interesting to know how the influence of that test played out in the classrooms.

Some insight into that question might be drawn from the observations of Ms. Drake's Fundamentals class. That class was specifically for eighth-grade students who

had not yet passed the FMT as of October of that school year. Among the six classes that were observed, Ms. Drake's Fundamentals class was the least aligned with reform instruction or any instruction that would be considered preparation for the PA. This class also received a reform score equal to the minimum reform score of all surveyed teachers. However, this is only one class, and it might have been an extreme example in more than one sense: This class was made even more homogeneous since its roster had been reassigned during the school year, and the teacher used an exceptionally non-reform instructional style with that class.

In summary, teacher credentials and teacher content knowledge inevitably have an impact on what teachers do in the classroom. Additionally, pressures from a high stakes test that many or all of their students will face could influence teachers' instructional style further. The quantitative analysis in this study showed that credentials do make a difference, and pressure from another test is certainly on the minds of teachers. Exactly how those influences manifest themselves in the learning environment would be substantive topics for future studies. The qualitative analysis was able to provide some insight into how limited content knowledge and pedagogical content knowledge might hamper construction of conceptual understanding in mathematics through, among other things, the teachers' inability to question effectively and cultivate productive discourse.

Predicting Achievement on a State Performance Assessment

This study was concerned with whether students in different tracks, or course levels, have the same opportunity to learn skills and concepts assessed on the State PA.

The third and final research question focused on the relationship between student achievement across courses and instructional practices employed by teachers. Initially, it was hypothesized that teachers might use certain strategies with higher ability students more than with lower ability students. Those strategies were thought to be more in line with reform recommendations, which are commensurate with goals of instruction outlined in the State's rationale for the PA. It was thought that such a relationship between level of course and type of instruction might be responsible for an achievement gap. However, this study found no links between course level and reform instruction, and furthermore, no explicit link between reform instruction and achievement on the PA, overall, although for Algebra II, more reform meant slightly more achievement.

Aside from the disappointing result that sixth-grade CTBS scores explained nearly all of the variation in PA achievement, teacher credentials appeared to be a culprit in the link between course level and achievement. As discussed earlier in this section, teachers who had less mathematics credentials were over-represented in the lower level courses, and their students performed significantly worse on the PA than did students in the other courses. The finding that credentials mattered above and beyond course level begs further research. Targeted areas for further study might be in the areas of teacher knowledge of mathematical content and pedagogy and the links to teacher practices and teacher questioning.

Limitations

There were some statistical limitations to this study, as well as other limitations that were more qualitative in nature. First, the sample may not necessarily include a representative group of all eighth-grade mathematics classes in this district. Although all teachers of eighth-grade mathematics were surveyed, the class about which they were surveyed was chosen in a way to maximize variation in courses sampled. For instance, if a school had three teachers, only one of whom taught Algebra II, that teacher was asked to answer the questionnaire considering his or her Algebra II class. The other two teachers would be asked to consider a particular class also with the same goal in mind. Thus, although that Algebra II teacher might have also taught Pre-Algebra in that school, he or she would not have been in the Pre-Algebra sample. Although this was not an enormous issue, it is important to state that classes for teachers were not chosen randomly from among their schedule.

Additionally, the results of this study should likely not be generalized to the state, as only one district was included in the study. Districts in this state can (and do) use different curricula, could advocate different lesson structures, emphasize different teaching approaches, or provide different amounts of professional development than the district studied here.

In this study, class level data was analyzed because of the teacher variables involved. That is, I asked teacher-level (classroom level) questions. Ideally, hierarchical linear modeling (HLM) would be used to take advantage of individual student scores, but also appropriately handle class-level, and even school-level variables. This approach was not possible for this study given the constraint of very few teachers within certain

schools. Nonetheless, HLM might be fruitful in future analyses as a way to address questions of how teacher instructional strategies affect variations in students' PA scores or whether certain strategies are more or less effective for students of varied abilities.

Another limiting aspect of studies like the one presented here is having only one measure of achievement with which to try to determine how different levels of classes performed following certain "treatments." In the Results section, I alluded to the notion that in light of the CTBS scores being so predictive of PA achievement, the gap between low course-level and high course-level students remained the same; actually, the methodology of this study did not allow for an examination of the gap itself. The surrogate for a pre-test score was a CTBS score, the only available data that measured mathematics prior knowledge. Pre-test/post-test measures or growth modeling would allow a researcher to study how students' mathematics ability or performance change over time. I bring this up, not as a limitation on this study, since those methods simply could not have applied in this case where a particular one-time, eighth-grade assessment was given. However, growth modeling has the potential to examine students' (a) individual growth over time and (b) how teacher instructional practices (or other predictors) affect variation in students' rate of growth (Willett, 1994). With these methods, the *gap* between achievers could be examined for change.

This study attempted to characterize what teachers do in their classrooms as either reform or traditional practice, and subsequently look at differentials in achievement. Other areas of research on reform instruction have focused on how students who learn *within a reform curriculum* fare in comparison to students whose curriculum and learning environment are more traditional. In those studies, the reform

curriculum, and accompanying trained teachers, essentially vouches for the status of the learning environment as being reform. Likewise, the absence of a reform-based curriculum is presumed to allow for the designation of the opposing learning environment as traditional. In the district in this study, although different courses were examined, the curricula were not targeted as either reform or traditional. Perhaps because of a lack of differently oriented curricula, and the fact that the district as a whole had certain expectations, such as the structure of a lesson, there was less variability in reported instructional strategy use overall. Teachers reported using most of the strategies for some amount of time. As a result, reform scores did not vary very much, and there might not have been enough power in the sample size to detect real differences from the narrowly varying scores among the three courses.

Although the classroom observations lent some credence to the questionnaire's ability to at least rank teachers on a scale of traditional to reform teaching, the observations also revealed the questionnaire's limitations. Self-report data is easily called into question. Other methods of differentiating mathematics classroom environments as either reform or traditional, such as using videotapes followed by an "expert's" rating, might be more effective, assuming the cost and time involved is not prohibitive. Other researchers argue for the importance of including all of these significant elements to assess the level of reform in a classroom environment: curriculum, teacher beliefs, teacher activities, and student activities (Calhoun, Bohlin, Bohlin, & Tracz, 1997), as opposed to focusing on only one, as this study and many others have done. However, the study by Calhoun et al. also found a high degree of correlation between all of those elements, with student activities being able to singularly

account for variability in classroom environment classification. The relevant questionnaire items in this study essentially asked teachers to report on student activities. Thus, the self-report issue notwithstanding, the questionnaire used in this study did focus on the most important element in assessing level of reform instruction.

Recommendations

Studies looking at reform curricula, reform teaching strategies, teacher beliefs aligned with reform, and student beliefs about mathematics have come out of the constructivist research agenda. Much of the early research in these areas was qualitative and often not directly concerned with student achievement. Accountability pressures have made it incumbent upon mathematics educators, who continue to make claims about the benefits of the reform movement, to link reform measures with measures of achievement, particularly in middle and secondary school, where it is most lacking.

This study was ultimately unable to provide evidence for a relation between teaching practice (reform or traditional) and achievement on a state performance assessment. It would be interesting to know if instructional approach within a given curriculum relate to student performance in classroom measures, such as unit tests, performance tasks, and grades. This study was also unable to identify differences in levels of reform teaching for different courses. However, the observational study revealed that the survey data likely failed to truly capture the nature of the classroom environment – the real level of student autonomy and initiative, as well as student dialogue throughout concept development, elements that are recommended in constructivist philosophy (Brooks & Brooks, 1993). The observational study also

suggests that teachers' limited content or pedagogical knowledge impedes their ability to foster environments where meaningful learning of mathematics can take place.

Future research efforts should focus on the connections between teacher content knowledge, pedagogical content knowledge, and learning environment, including but not limited to instructional strategies. The nature and role of discourse in mathematics classes and how it influences student learning of concepts is an essential piece of the mathematics education puzzle. Measuring that learning through a variety of assessments, including high-stakes assessments, and linking the achievement back to the learning environment must be part of the research agenda. Ideally, if the same assessment measure is used, growth modeling could be employed to follow students longitudinally.

One final comment is the need to make sure that students in all classes have teachers appropriately credentialed in mathematics. If this expectation is unrealistic – if a main tenet of the No Child Left Behind Legislation fails – research regarding teacher credentials and its influence on the learning environment must be investigated.

APPENDIX A: LETTER TO PRINCIPALS

[Date]

Dear [Principal's name],

Please allow me to introduce myself. I am a mathematics teacher at Baltimore City Community College, as well as a graduate student in mathematics education at the University of Maryland. I am writing to request your cooperation in my study of *eighth-grade mathematics teachers' instructional practices across tracks and the relationship of those instructional practices to student performance on the mathematics component of [the PA]*. I have already obtained approval from [the district's research director] in the Research Office (letter attached), and have the full support of the Office of Mathematics for doing this research. I am now asking you to consider permitting your school's participation in this study.

As you know, the mathematics eighth-grade [State PA] is not a traditional mathematics achievement test. In order for students to do well on the [PA], many educators believe that different instructional practices in mathematics are necessary. The purpose of my study is to ascertain what kinds of instructional practices are being used by eighth-grade mathematics teachers, and whether there is a corresponding relationship to achievement on the [PA].

Let me briefly outline what the study would entail for personnel at your school, and the time frame involved. My study surveys eighth-grade mathematics teachers' opinions and practices regarding mathematics instruction. Teachers will think about a particular section or class that they teach when answering the survey. The survey will be administered in two short sections, since certain questions can be more accurately answered toward the end of the year. The first section is a two-page, two-sided questionnaire and will be sent in November; the other section is a shorter survey, and will be sent in March. Two dollars will be attached to each questionnaire on both occasions as a token of my appreciation, although *it will be made clear that completing the survey is strictly voluntary*. Both times, I will enclose a stamped, return envelope so the questionnaire can be returned to me directly. [District] offices, participating teachers and cooperating principals who are interested in the final study will receive data compiled in statistical summary form only. Individual teachers' responses will be kept strictly confidential. Individual surveys will be coded with an identification number, not by name. No individual responses or surveys will be released to district personnel. I need approximately 80 eighth-grade mathematics teachers to complete the survey.

A second aspect of my study involves looking at the classroom practices of two eighth-grade mathematics teachers from [your county] identified as exemplary by their principals or mathematics content leaders. Thus, in January I will send a letter to you

asking if there are any eighth-grade mathematics teachers whom you believe to be outstanding with respect to particular teaching styles, and whom you would recommend for this study. If you or the mathematics content leader choose to identify a teacher from your school, I would approach that teacher, and ask him or her if they would agree to my visiting his or her classroom several times over a period of about one month (sometime between January and early April). Of course, you may choose not to identify a teacher or a teacher that you identify may choose not to participate. The names of cooperating teachers will not be released; the purpose of these visits is not to evaluate the teachers, but rather to describe exemplary classroom practices. The two selected teachers will set the dates so that the visitations are at their convenience.

I will speak briefly to your school's mathematics content leader prior to distributing the initial survey. This is so I will better understand the characteristics of the eighth-grade mathematics program offered at your school. This conversation need not be in person. I will simply speak over the phone with the mathematics content leader at his or her convenience during September or October. I will be asking the mathematics content leader for the names of the eighth-grade mathematics teachers in your building along with their schedule for the year, so that I can identify the class period that the teacher should think of when completing the survey.

The [State PA] data needed for my study will be supplied by the [District Central Office]. This data will be in class-aggregated form so that student confidentiality laws are respected.

I hope that you will give my request careful consideration. To summarize, I will talk to the mathematics content leader in your school and survey each eighth-grade mathematics teacher. Then there may be an exemplary teacher from your school selected for classroom visits. Ultimately, each teacher will have the option of participating to the extent they choose (survey and visits, survey only, no participation). I would appreciate your support tremendously. I will call you to discuss this request very shortly and will address any concerns you may have at that time. If you wish to get in touch with me at any time, please call me at 410-542-8372. Thank you for your cooperation.

Sincerely,

Felice Shore

APPENDIX B: SURVEY, PART 1

APPENDIX C: SURVEY PART 2

[Last page of Survey, Part 2 to Algebra I and II teachers]

APPENDIX D: LETTER TO PRINCIPALS FOR OBSERVATIONS

[Date]

[School address]

Dear [Principal's name]

First, I'd like to thank you for your school's participation in my study of eighth-grade mathematics instruction and its relationship to achievement on the [PA]. In the fall, I surveyed eighth-grade mathematics teachers in 20 middle schools in [the district] (including yours) and attained over 80% response. Additionally, many teachers wrote notes of encouragement and support. I am grateful for your support.

As I explained in the September letter, I am a mathematics teacher at Baltimore City Community College and also a graduate student at the University of Maryland. The study I am conducting is for a dissertation in mathematics education. In the previous letter to you I described a second component to my study that involves documenting the kinds of things that outstanding teachers do in their classrooms. This is where I could use your help. If you have an eighth-grade teacher of mathematics whom you consider to be exemplary, I would very much appreciate it if you would let me know about him or her. I would then contact that teacher, explain the purpose of the study, and then *ask* if (s)he would agree to my observing a selected class over a period of several weeks, though not every day. Of course, these observations would be strictly confidential.

My goal is to identify two such teachers in [the district] with whom to carry out a thoughtful qualitative analysis of classroom instruction. I would like the classroom visits to occur between February and early April at the teacher's convenience.

Again, if you have any teachers whom you believe to be especially effective with eighth-grade mathematics students, please call me at 410-542-8372. Please realize that even if you recommend a teacher, it is still up to that teacher to say whether (s)he is willing to participate in this part of the study. I hope to hear from you as soon as possible. Thank you very much.

Sincerely,

Felice Shore

APPENDIX F: POST-OBSERVATION NOTES

Post-observation: Reflection / Interpretation

(Give detailed answers to the following questions as soon after the observation as possible)

1. What kind of mathematical activity were students engaged in (and for how much time)?
2. What was the nature of the questions the teacher asked of students (procedural / conceptual; extension / review)
3. How did the teacher encourage students to explain/clarify to the class or each other?
4. Were there indications of the teacher's attempts to understand student meanings or consider student conceptualizations?
5. Was content/ were concepts presented or developed? How? By students or teacher?
6. Describe efforts to support classroom interaction / investigation / mathematical understanding.
7. How did the teacher make an attempt to involve **all** students?
8. What aspects of today's class might help prepare students for the state performance assessment?

REFERENCES

- Artzt, A.F., & Armour-Thomas, E. (1999). A cognitive model for examining teachers' instructional practice in mathematics: A guide for facilitating teacher reflection. *Educational Studies in Mathematics, 40*, 211-235.
- Ball, D. L. (1996). Teacher learning and the mathematics reforms. What we think we know and what we need to learn. *Phi Delta Kappan, 77*, 500-507.
- Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Battista, M. T. (1994). Teacher beliefs and the reform movement in mathematics education [Electronic version]. *Phi Delta Kappan, 75*, 462-467.
- Baxter, J. A., Woodward, J., & Olson, D. (2001). Effects of reform-based mathematics instruction on low achievers in five third-grade classrooms [Electronic version]. *Elementary School Journal, 101*, 529-547.
- Bay, J. M. (2000, April). The dynamics of implementing and sustaining standards-based mathematics curricula in middle schools. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education, 33*, 239-258.
- Bond, L., Roeber, E., & Connealy, S. (1998). *Trends in state student assessment programs*. Washington, DC: Council of Chief State School Officers.
- Bottage, B. (1999). Effects of contextualized mathematics instruction on problem solving of average and below-average achieving students. *Journal of Special Education, 33* (2), 81-92.
- Bracey, G. W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappan, 68*, 683-686.
- Brooks, J. G., & Brooks, M. G. (1993). *The case for constructivist classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Calhoun, D. O., Bohlin, C., Bohlin, R., & Tracz, S. (1997, March). *The mathematics reform movement: Assessing the degree of reform in secondary mathematics classrooms*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education, 11*, 81-95.
- Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C.P., & Loeff, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*, 499-531.
- Cobb, P. (1994). Constructivism in mathematics and science education. *Educational Researcher, 23*(7), 4.
- Cobb, P., Wood, T., & Yackel, E. (1990). Classrooms as learning environments for teachers and researchers. *Journal for Research in Mathematics Education, 4*, 125-146.
- Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22*, 3-29.
- Cohen, D. K. (1995). What is the system in systemic reform? *Educational Researcher, 24*(9), 11-17, 31.
- Cohen, D. K., & Ball, D. L. (1990). Relations between policy and practice: A commentary. *Educational Evaluation and Policy Analysis, 12*, 249-256.
- Creswell, J.W. (1994). *Research design*. Thousand Oaks, California: SAGE.
- Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438-481.
- Darling-Hammond, L. (2000). New standards and old inequalities: School reform and the education of African American students. *Journal of Negro Education, 69*, 263-287.
- Doyle, W. (1983). Academic Work. *Review of Educational Research, 53*, 159-199.
- Driver, R. (1987). Promoting conceptual change in classroom settings: The experiences of the children's learning science project. In J. Novak (Ed.), *Proceedings of the Second International Seminar. Misconceptions and Educational Strategies in Science and Mathematics Education*. (Vol. 2, pp.97-107). Ithaca, NY: Cornell University.

- Driver, R. (1995). Constructivist approaches to science teaching. In L. P. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 385-400), Hillsdale, NJ: Erlbaum.
- Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*, 95-113.
- Firestone, W.A., Winter, J., & Fitz, J. (2000). Different assessments, common practice? Mathematics testing and teaching in the USA and England and Wales [Electronic version]. *Assessment in Education: Principles, Policy & Practice, 7* (1), 13-37.
- Fowler, F. J. (1993). *Survey research methods*. Newbury Park, CA: SAGE.
- Gall, M. (1984). Synthesis of research on teachers' questioning. *Educational Leadership, 42* (3), 40-47.
- Gamoran, A. (1992). Is ability grouping equitable? *Educational Leadership, 50* (2), 11-17.
- Gamoran, A. (1995). An organizational analysis of the effects of ability grouping. *American Education Research Journal, 32*, 687-715.
- Ginsburg-Block, M. D., & Fantuzzo, J. W. (1998). An evaluation of the relative effectiveness of NCTM Standards-based interventions for low-achieving urban elementary students [Electronic version]. *Journal of Educational Psychology, 90*, 560-569.
- Graue, M. E., & Smith, S. Z. (1996). Parents and mathematics education reform: Voicing the authority of assessment. *Urban Education, 30*, 395-421.
- Gregg, J. (1995). The tensions and contradictions of the school mathematics tradition. *Journal for Research in Mathematics Education, 26*, 442-466.
- Guskey, T.R. (1994). What you assess may not be what you get. *Educational Leadership, 51* (6), 51-54.
- Haberman, M. (1991). The pedagogy of poverty. *Phi Delta Kappan, 73*, 290-294.
- Hancock, G. R., & Klockars, A. J. (1996). The quest for α : Developments in multiple comparison procedures in the quarter century since Games (1971). *Review of Educational Research, 66*, 269-306.
- Hawk, P.P., Coble, C.R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education, 36* (3), 13-15.

- Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12 (4), 20-25, 41-42.
- Herman, J. L., Klein, D. C. D., Heath, T. M., & Wakai, S. T. (1994). *A first look: Are claims for alternative assessment holding up?* (Tech. Rep. No. 391). Los Angeles: University of California, Center for the Study of Evaluation, Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Hewson, P. (1996). Teaching for conceptual change. In D. Treagust, D. Reinders, & B. Fraser (Eds.), *Improving teaching and learning in science and mathematics* (pp. 131-140). New York: Teachers College Press.
- Hiebert, J., & Carpenter, T.P.(1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). New York, NY: Macmillan.
- Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis*, 14, 205-227.
- Huntley, M., Rasmussen, C. L., Villarubi, R. S., Sangtong, J., & Fey, J. T. (2000). Effects of Standards-based mathematics education: A study of the Core-Plus Mathematics Project. *Journal for Research in Mathematics Education*, 31, 328-361.
- Jacobs, J. K., & Morita, E. (2002). Japanese and American teachers' evaluations of videotaped mathematics lessons. *Journal of Research in Mathematics Education*, 33, 154-175.
- Kerckhoff, A.C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review*, 51, 842-858.
- Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996). *Final Report: Perceived effects of the Maryland School Performance Assessment Program* (Tech. Rep. No. 409). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Larson, R. E., Kanold, T. D., & Stiff, L. (1993). *Algebra 2: Teacher's Edition*. Lexington, MA: D.C. Heath.
- Linchevski, L., & Kutscher, B. (1998). Tell me with whom you're learning, and I'll tell you how much you've learned: Mixed ability versus same-ability grouping in mathematics. *Journal for Research in Mathematics Education*, 29, 533-554.
- Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, 15, 1-16.

- Lubienski, S. T. (2000). Problem solving as a means toward mathematics for all: An exploratory look through a class lens. *Journal for Research in Mathematics Education*, 31, 454-482.
- Mathematics Plus*. (1992). Orlando, FL: Harcourt Brace Jovanovich.
- Mathison, S. (1991). Implementing curricular change through state-mandated testing: Ethical issues. *Journal of Curriculum and Supervision*, 6, 201-212.
- Mayer, D. (1998). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis*, 20, 53-73.
- Mayer, D. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, 21, 29-45.
- Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11 (1), 3-9.
- Miller, M. D., & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*, 12, 9-15.
- The Maryland School Performance Assessment Program (MSPAP)* (1998). Baltimore: Maryland State Department of Education.
- National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal*, 84, 113-130.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (1991). *Professional Standards for teaching mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academy Press.
- Nobel, A. J., & Smith, M. L. (1994). Old and new beliefs about measurement-driven reform: "Build it and they will come." *Educational Policy*, 8, 111-136.

- Oakes, J. (1986). Beyond tracking. *Educational Horizons*, 65 (1), 32-35.
- Oakes, J. (1991). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: RAND.
- Peterson, J. M. (1989). Remediation is no remedy. *Educational Leadership*, 46 (6), 24-25.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679-682.
- Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal*, 30, 523-553.
- Resnick, L. B. (1981). Testing in America: A supportive environment. *Phi Delta Kappan*, 62, 625-628.
- Reys, R., Reys, B., Barnes, D., Beem, J., & Papick, I. (1998). What is standing in the way of middle school mathematics curriculum reform? *Middle School Journal*, 30 (2), 42-48.
- Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of Standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, 34, 74-95.
- Rickard, A. (1995). Teaching with problem-oriented curricula: A case study of middle-school mathematics instruction. *The Journal of Experimental Education*, 64 (1), 5-26.
- Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education*, 32, 368-398.
- Robinson, S. P. (1996). With numeracy for all: Urban schools and the reform of mathematics education. *Urban Education*, 30, 379-394.
- Romberg, T.A., Wilson, L., Khaketla, M., & Chavarria, S. (1992). Curriculum and Test Alignment. In T. A. Romberg (Ed.), *Mathematics assessment and evaluation: Imperatives for mathematics educators* (pp.61-74). Albany: State University of New York Press.

- Schoen, H. L., Finn, K. F., Griffin, S. F., & Fi, C. (2001, April). Teacher variables that relate to student achievement in a standards-oriented curriculum. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Shepard, L. A. (1991). Will national tests improve student learning? *Phi Delta Kappan*, 72, 232-238.
- Shepard, L. A. (1995). Using assessment to improve learning. *Educational Leadership*, 52(5), 38-43.
- Shepard, L. A., & Dougherty, K. C. (1991, April). *Effects of high-stakes testing on instruction*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Silver, E. A. & Stein, M. K. (1996). The QUASAR Project: The “Revolution of the possible” in mathematics instructional reform in urban middle schools. *Urban Education*, 30, 476-521.
- Simon, M. A., & Schifter, D. (1993). Toward a constructivist perspective: The impact of a mathematics teacher inservice program on students. *Educational Studies in Mathematics*, 25, 331-340.
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60, 471-499.
- Smith, J. (1996). Does an extra year make a difference? The impact of early access to algebra on long-term gains in mathematics attainment. *Educational Evaluation and Policy Analysis*, 18, 141-153.
- Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10, 7-11.
- Spillane, J.P. & Zeuli, J.S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21, 1-27.
- Stake, R. E. (1995). The invalidity of standardized testing for measuring mathematics achievement. In T. A. Romberg (Ed.), *Reform in school mathematics and authentic assessment* (pp. 173-235). Albany: State University of New York Press.

- Stecher, B.M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing. (CSE Tech. Rep. No. TR-482). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematics tasks used in reform classrooms. *American Educational Research Journal*, *33*, 455-488.
- Strauss, R.P., & Sawyer, E.A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, *5* (1), 41-48.
- Underhill, R (1988). Focus on research into practice in diagnostic and prescriptive mathematics. Mathematics learners' beliefs: A review. *Focus on Learning Problems in Mathematics*, *10*(1), 55-69.
- United States Department of Education. National Center for Educational Statistics. (1999). *What happens in classrooms? Instructional practices in elementary and secondary schools, 1994-1995* (NCES 1999-348). Washington, DC: Author.
- Wiley, D. E., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System (CLAS). *Educational Evaluation and Policy Analysis*, *17*, 355-370.
- Willett, J. B. (1994). Measuring change more effectively by modeling individual growth over time. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education*, (2nd ed., pp. 671-678). Oxford, UK: Pergamon Press.
- Wilson, S. M. (1990). A conflict of interest: Constraints that affect teaching and change. *Educational Evaluation and Policy Analysis*, *12*, 309-326.
- Worthen, B. R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, *74*, 444-454.